W.E. UPJOHN INSTITUTE FOR EMPLOYMENT RESEARCH
Researching the causes and consequences of unemployment

Pre-K in the Public Schools: Evidence from within U.S. States

Upjohn Institute Working Paper 18-285

Timothy J. Bartik (bartik@upjohn.org)
Brad Hershbein (hershbein@upjohn.org)

Corresponding author: Timothy J. Bartik (bartik@upjohn.org)
Phone number: (+1) (269) 385-0433

Affiliation for both Bartik and Hershbein:
W.E. Upjohn Institute for Employment Research
300 S. Westnedge Avenue
Kalamazoo, MI United States 49007

May 1, 2018

ABSTRACT

In the past 15 years, four-year-olds' enrollment in state-funded pre-kindergarten in the United States has doubled, and advocates have pushed for further expansion. Although research has shown that pre-K programs can have important benefits, most existing studies have focused on small or state-specific programs that may not generalize to other areas or contexts. The uniqueness of our paper is its scope: our data cover the last two decades, span nearly all states, and allow for intrastate variation. For the average state program, we find no evidence of effects on the average student's test scores, assignment to special education, or grade retention. Our estimates rule out pre-K impacts as small as 2 percentiles. However, these averages conceal some important heterogeneity. In states previously found to have high-quality pre-K, we find positive effects on math test scores. For majority-black districts, the average pre-K program has large effects on math and reading.

Keywords: Pre-K, early childhood education, NAEP, test scores, sleeper effects

JEL Codes: H75, I21, I24, I28

Acknowledgments:
This work was supported by the Russell Sage Foundation (grant number 83-14-20). However, the Russell Sage Foundation was not involved in the study design, in the collection, analysis and interpretation of data, in the writing of the report, or in the decision to submit this article for publication; these tasks are solely attributable to the authors. We thank the Russell Sage Foundation for its generous support. We thank Matthew Chingos, Chloe Gibbs, and participants of the 2016 Western Economic Association meetings, the 2017 University of Wisconsin Institute for Research on Poverty workshop, and the 2018 Association for Education Finance and Policy meetings for helpful comments. Wei-Jang Huang and Nathan Sotherland provided invaluable research assistance. All findings are our own and should not be construed as reflecting the views of the Russell Sage Foundation or the Upjohn Institute.

Does the average state or local pre-K program provide former participants with persistent benefits? If not, do high-quality pre-K programs do so? If benefits are provided, do they occur for all students, or only for disadvantaged groups? Despite considerable research evidence on the outcomes of pre-K, the answers to these questions are still controversial and unsettled. The current paper provides new evidence on these questions from a unique linked database that covers almost all states and over two decades of public pre-K programs.

For some early childhood programs that were run a long time ago, such as Perry Preschool, the Abecedarian Project, and the Chicago Child-Parent Center (CPC) program, the research evidence is strong: these pre-K programs produce sizable benefits for former participants, both in the short-term and long-term. The findings of sizable pre-K benefits from these programs has helped generate support from the public and policymakers for significant expansions of public pre-K. Indeed, state-funded pre-K programs have grown from covering 14 percent of all four-year-olds in 2001–2002 to 32 percent in 2015–2016 (Barnett et al. 2017). While some of these pre-K program expansions have been targeted at disadvantaged children, others have been universal, open to all students regardless of disadvantaged status. For example, Oklahoma has since the early 2000s run a near-universal state pre-K program, currently enrolling 74 percent of the state's four-year-olds. More recently, New York Mayor Bill de Blasio campaigned on and implemented a universal pre-K program.

However, critics have raised doubts about whether the strong results from the early public programs are likely to apply to these larger and sometimes universal public pre-K programs (Stevens and English 2016). The Perry, Abecedarian, and Chicago programs were all high-quality and expensive programs run on a relatively small scale, and targeted on disadvantaged students. The average state program may not be as high-quality, and certainly is not as expensive per student. Furthermore, pre-K may have smaller effects for more advantaged students, who typically have greater family and neighborhood resources—and private pre-K options—during early childhood. More recently, evidence from Tennessee has been used to argue that the state's pre-K program has positive effects only at kindergarten entrance, with test score gains quickly disappearing by early elementary school (Lipsey, Farran, and Hofer 2015). On the other hand, others have pointed to evidence from North Carolina and Oklahoma to argue that these two states' programs have persistent benefits, through at least 4th grade for North Carolina (Ladd, Muschkin, and Dodge 2014) and through 8th grade in Oklahoma (Bartik et al. 2016). Other studies have looked at Georgia (sometimes in conjunction with Oklahoma) and found some evidence for persistent benefits, at least for disadvantaged students (Cascio and Schanzenbach 2013; Fitzpatrick 2008).

Notably, these studies of large-scale state programs are typically of only one or two states. Very few studies examine pre-K programs throughout the country, and it is unclear whether the findings from local or even state programs generalize. Additionally, many of the specific programs studied may be of unusually high "quality" (a rationale for studying them in the first place) and may not reflect the typical pre-K programs that have been, and continue to be, implemented. Furthermore, studies of specific statewide programs that use cross-state comparisons may not adequately control for other factors that could influence educational achievement, such as other state policies, and impact estimates are often relatively noisy, leading to uncertainty about true effects. The empirical evidence on the effectiveness of typical state pre-K programs is thus limited.

In this paper we perform the first national analysis of public pre-K participation on standardized test scores, special education assignment, and grade retention using within-state variation. We match detailed, student-level microdata from the National Assessment of Educational Progress (NAEP), the Nation's Report Card, to public pre-K enrollment at the school district level for different types of students and districts.[1] For the 4th graders taking the NAEP, we use the Common Core of Data from the National Center for Education Statistics to estimate their likelihood of being enrolled in pre-K five years ago. Our data stretch from pre-K enrollments in the early 1990s (4th grade outcomes in the mid-to-late 1990s) through pre-K enrollments in 2008 (4th grade outcomes in 2013), offering substantial variation in public pre-K across time and space.

To identify the impact of pre-K on student outcomes, we adopt a two-stage augmented differences-in-differences methodology. The first stage uses student-level data in NAEP to calculate means at the geography-year cell net of individual student characteristics. The second stage takes these collapsed, adjusted means and implements a differences-in-differences specification controlling for geography and time fixed effects, and sometimes higher-level interactions. The extent of pre-K variation allows for more precise estimates than most previous studies, although it comes at the expense of program specificity. That is, instead of estimating the effect of a specific pre-K program on later outcomes, we effectively estimate the "average" effect of pre-K diffusion through public schools on both academic and nonacademic outcomes. The

[1] As explained later, we also report for comparison results at the state level, and, consistent with previous studies, find that such results are very imprecise. In sensitivity tests in the appendix, we also consider effects at the 8th grade level and find results consistent with our 4th grade estimates, although there are necessarily fewer observations and more imprecision with such longer-term follow-up. Finally, the sensitivity tests also consider effects of pre-K at the school level, but we take these results less seriously because of the higher student mobility across schools.

data allow us to estimate effects for students overall as well as for different groups of students (or districts), stratified by race, income, and other characteristics.

Our overall findings are that the average public pre-K program has statistically and substantively insignificant effects on 4th grade outcomes. Our estimates are generally precise enough to rule out student outcome effects from full pre-K adoption of 2 percentiles in math and reading test scores and 3 percentage points in special education assignment and grade retention. The magnitude and precision of our estimates rule out a very high rate of return to the typical pre-K program in a social benefit-cost analysis.

However, we find that the null result for the average masks important heterogeneity, particularly for districts in states with high-quality programs and those with a high black student membership.[2] Based on prior research, we classify five states as having high-quality pre-K programs: Maryland, Massachusetts, New Jersey, North Carolina, and Oklahoma. In these five states, shifting from no public pre-K to full adoption of public pre-K improves 4th grade math test scores by a statistically significant 2.8 percentiles, about twice the necessary magnitude to pass a benefit-cost test in terms of predicted future earnings increases. For districts that are majority black, the point estimates suggest large test score benefits of 5.9 percentiles in math and 3.8 percentiles in reading. Among the five "quality states," estimated effects in majority-black districts are larger still, at 6.6 percentiles for math and 7.4 percentiles for reading. In contrast with the average public pre-K program, the point estimates for programs in quality states and majority-black districts suggest sizable benefit-cost ratios.

The remainder of this paper is organized as follows. The next section outlines the conceptual and methodological challenges in estimating the impact of pre-K. In the context of

[2] In an appendix, we also explore how results vary with district size and with district percent of students eligible for free or reduced-price lunch.

these challenges, we review and interpret findings from the large and growing pre-K research literature. We then describe our empirical approach. The conclusion presents our results, places them in the context of existing literature, and discusses possible future research directions.

THE RESEARCH LITERATURE ON PRE-K: IMPLICATIONS FOR THIS STUDY

In this section, we briefly review the large research literature on pre-K. Our summary focuses on the research findings and limitations that are most relevant to our current study. Appendix Table A1 provides a more detailed listing of results from most of the prominent pre-K studies, including estimated pre-K effects at the short-, medium-, and long-terms. We identify seven aspects of the literature important to understanding the present study.

1) Only modest medium-term effects are necessary for pre-K to have predicted long-term benefits greater than costs. The average state pre-K program costs $5,696 per student (National Institute for Early Education Research [NIEER] 2017).[3] Based on the estimated relationship between elementary school test scores and adult earnings (Chetty et al. 2011), 4th grade test scores would need to improve by about 1.4 percentiles to yield an earnings increase of $5,696 in present value terms.[4] Special education is sufficiently expensive that pre-K would need to reduce such assignment by only 1 student in 28 (assuming assignment is permanent through

[3] This figure is for the school year 2015–2016, attempts to include both state and local spending, and averages the costs of half-day and full-day programs using the existing mix of half-day and full-day students.

[4] Chetty et al.'s (2011) research suggests that at 4th grade, each 1 percentile increase in test scores increases future earnings by roughly 0.5 percent of mean overall earnings. The present value, discounted back to age 4, of future mean overall earnings in the United States is around $817,000 (in 2016 dollars; Bartik [2014]). Therefore, an increase in 4th grade test scores of 1.39 percentiles is sufficient to increase expected future earnings by $5,696: 1.39 percentiles = $5,696 / ($817,000 × 0.005). For at least some pre-K programs for the disadvantaged, the benefits from crime reduction might be of the same order of magnitude as earnings benefits (Bartik et al. 2016), which implies that the cutoff percentile for such pre-K programs for the disadvantaged could be even less, about 0.70 percentiles. However, for universal programs, crime reduction benefits are likely much smaller, as average baseline crime rates are also much smaller (Bartik et al. 2016).

high school, which is common) to pay for itself.[5] The costs of retention in grade through future earnings losses and increased crime are estimated to be $195,439 per retained student in present value, in 2016 dollars (Bartik et al. 2016). Therefore, the costs of pre-K are recouped if the program can reduce grade retention by 2.9 percentage points.[6]

2) Select pre-K programs can have very large effects, average pre-K programs often have more moderate effects. Several experimental studies, and many good nonexperimental studies, find large short-term and long-term effects of pre-K on student outcomes. However, these studies by necessity are limited to selected programs—often higher quality programs—and may not apply to average state and local public pre-K programs, which are what we examine in this study. Two classic experimental studies from the 1960s and 1970s, those of the Perry Preschool and the Abecedarian program, found large and enduring effects on former participants' outcomes. Adult earnings, for example, were 19 percent higher in Perry and 26 percent higher in Abecedarian. Short-term effects (e.g., at the end of pre-K or beginning of kindergarten) in both studies included an increase in test scores of almost 20 percentiles.[7] However, Perry Preschool and the Abecedarian program are far more intense than the typical modern pre-K program, with Perry costing over $21,000 per student (in 2016 dollars) and Abecedarian over $89,000.

Studies of more recent (and slightly more typical) pre-K programs have also found short- and long-term effects, but of perhaps one-third to two-thirds those found for Perry and Abecedarian. A quasi-experimental study of Head Start estimated early test score gains of 5

[5] Special education costs are roughly 90 percent greater than regular education costs. According to the Common Core of Data, described below, average education costs in 2013–2014 were $13,470 (in 2016 dollars). Assuming that special education costs grow in tandem with the discount rate, the real cost of 13 years of special education would be $13,470 × 0.9 × 13 ≈ $157,600, or about 28 times the average pre-K costs of $5,696.

[6] $5,696 / $195,439 = 2.9 percentage points.

[7] Appendix Table A1 also shows impacts in the more-usual effect-size units.

percentiles and later effects that suggest a Head Start earnings boost of 11 percent (Deming 2009). Studies of the Chicago Child-Parent Study estimate early test score gains of 11 percentiles and adult earnings effects of 8 percent. Summarizing many similar studies, meta-analyses of the pre-K literature find immediate test score gains that average 9 to 14 percentiles (Camilli et al. 2010; Duncan and Magnuson 2013).[8] However, even though these studied programs are often closer to typical state pre-K programs, most are probably still of higher quality than is typical. For example, the oft-cited pre-K programs in Chicago and Tulsa both spent over $5,000 per student annually for a half-day pre-K program, considerably more than most equivalent-length state and local pre-K programs. The highly lauded (and full-day) pre-K program in Boston Public Schools spends over $15,000 per student annually.

Whether the results from these relatively high-cost programs generalize to the average or typical program is of current policy interest, and we try to address this question in the current study.

3) Gains in test scores from pre-K often fade substantially by late elementary school.

Many studies in the pre-K literature find extensive, if not total, fade-out of test score gains between kindergarten and middle (3rd through 8th) grades. In the meta-analyses of pre-K studies, test score effects decline by one-half to two-thirds over this horizon, with average middle-grade test score effects of 4 to 5 percentiles. Some studies find more complete fading: Chicago CPC, Head Start, Tennessee, and Perry. Nonetheless, the general pattern of results suggests that, if the typical public pre-K program has effects comparable to those from programs

[8] The age-based regression discontinuity (RD) design studies in Tulsa, Boston, and Tennessee (see Appendix Table A1) find somewhat larger effects. This may reflect that the regression discontinuity studies compare pre-K graduates with a control group that is further away in age from entering kindergarten, and therefore less likely to have attended pre-K the previous year, than is true of comparison groups in other pre-K studies. Thus, the counterfactual in RD pre-K studies involves students who have less education.

previously studied, it should be possible to detect them in the 4th grade, as we attempt here. As we will show, our district data allow sufficient precision to detect test score effects in most specifications that are much smaller than 4 percentiles.

4) **"Sleeper" effects of pre-K can reemerge later in life.** While many pre-K programs evince effect fade-out in middle grades, large effects often are found later in adulthood. This pattern is pronounced in the Perry program, the Chicago CPC program, Deming's (2009) study of the Head Start program, and Chetty et al.'s (2011) study of the effects of higher "kindergarten class quality." Some have argued that these sleeper effects may be due to program effects on soft or interpersonal skills (Heckman 2015; Heckman et al. 2013), which are difficult to measure (Duckworth and Yeager 2015). If soft skills are important to long-term effects, then it is important to try to measure the impact of pre-K on outcomes more tightly correlated with soft skills than standardized test scores. In the current study, we try to do so by examining pre-K's effects on grade retention and assignment to special education status at 4th grade.[9]

5) **There are conceptual differences between studies of pre-K participation at the individual student level and those of access based on geography.** While many of the abovementioned studies measure and examine an individual's actual participation in a pre-K program, other studies proxy participation through students' access to pre-K based on where they live. These latter, geographic studies sometimes find large test score effects several years after pre-K. For example, Cascio and Schanzenbach (2013) compare pre-K adoption in Georgia and Oklahoma with other states and find 4th grade test score effects of 14 percentiles. Ladd, Muschkin, and Dodge (2014) compare counties with different pre-K access in North Carolina

[9] Another longer-term, related behavioral outcome is the high school graduation rate, which we plan to analyze in the future.

and find 3rd grade test score effects of 20 to 25 percentiles.[10] As they point out, if "there were no spillover effects of the program to other children, the test score impacts would be unrealistically large." But such spillover effects are plausible, given evidence of peer effects found in Hanushek et al. (2003) and Hoxby (2000), as well as direct evidence found in Neidell and Waldfogel (2010) for positive kindergarten spillovers due to more students having attended pre-K. If such spillovers are present for the typical pre-K program, it would suggest a greater likelihood of finding positive impacts in the current study, which also uses a geographic access design.

6) Measuring pre-K quality is difficult. Although nearly all researchers agree program quality is important, there is little consensus on how to measure it. In most cases, there are rather weak relationships between existing structural measures of pre-K program quality (e.g., teacher credentials, class size, written curriculum, classroom physical features) and student learning (Bartik 2011, pp. 135–140; Sabol et al. 2013; Zaslow et al. 2010). Furthermore, observational measures of pre-K quality (e.g., trained observers attempting to objectively rate teacher-student interactions) also have only modest correlations with measures of pre-K learning. Some studies have found moderate positive correlations between CLASS (Classroom Assessment Scoring System) quality ratings and student learning (Keys et al. 2013), but not for other observational rating systems. However, other studies have found that higher CLASS ratings do not always predict better student outcomes (Burchinal, Kainz, and Cai 2011; Weiland and Yoshikawa 2013). Overall, this line of research suggests that "currently available quality measures may not be adequate to the research tasks being undertaken" (Keys et al. 2013). Nonetheless, because of the

[10] Although these large medium-run effects occur for Cascio and Schanzenbach (2013) and Ladd, Muschkin, and Dodge (2014), medium-run effects are much smaller (and sensitive to specification) in Fitzpatrick (2008) and Rosinsky (2014). The results in Cascio and Schanzenbach and Fitzpatrick are relatively imprecise as a consequence of having only one or two "treatment" states. When the standard errors are adjusted as suggested by Conley and Taber (2011), they become large enough that one can reject neither zero pre-K effects or very large pre-K effects.

recognized importance of quality, we attempt in the current study to examine how pre-K quality matters. Since it is difficult to quantitatively measure, we instead measure quality based on outside expert opinion.

7) **Pre-K relative to what? The counterfactual is important.** The estimated impact of pre-K can vary greatly depending on the counterfactual to a program. Indeed, different counterfactuals affect the interpretation of the results of the Head Start Impact Study, in which almost half of the randomly assigned control group attended some other early childhood program. Two recent papers show that the effect of Head Start relative to a counterfactual of no preschool are about 60 percent greater than its effect relative to a counterfactual that includes considerable preschool enrollment (Feller et al. 2014; Kline and Walters 2015).[11] Another recent paper shows how the diffusion of the television show *Sesame Street* in the late 1960s and early 1970s essentially functioned as an early childhood education program and improved schooling outcomes, in part because few children at the time were exposed to educational programming before elementary school (Kearney and Levine 2015). In the present study, we attempt to account for a counterfactual that includes alternate early childhood education program by controlling for the availability of Head Start and private preschool slots geographically near the public school district. Ideally, we would also control for the quality of those options, which we cannot do currently.

[11] As mentioned in a previous footnote, a different counterfactual may also help explain the generally greater short-term test score effects found in regression discontinuity studies of pre-K.

DATA AND METHODOLOGY

Data

Our data come from two main sources: the National Assessment of Educational Progress (NAEP), also called the Nation's Report Card, and the Common Core of Data (CCD).[12] Both data sets are maintained by the U.S. Department of Education. We also rely on expert opinion to indicate whether a state has a high-quality pre-K program.

National Assessment of Educational Progress

The NAEP is a nationally representative standardized assessment of students in certain academic subjects and grades, and it is the only uniformly administered test that is comparable across states and time.[13] The core subjects of mathematics and reading are currently tested biennially, in odd-numbered years, for the 4th and 8th grades. In this paper, we focus on the 4th grade.[14] Since 2003, every state has participated in the core NAEP tests, and the large sample sizes—approximately 3,000 students per state for each test administration in grade 4—are sufficient to allow for detailed analyses of student groups. Prior to 2003, the math and reading tests for grade 4 were administered less frequently, about every four years, with participation by most but not all states.

NAEP data at the state level are publicly available (https://nces.ed.gov/nationsreportcard/naepdata/dataset.aspx) and have been used in previous analyses of the effect of pre-K programs on student achievement (Cascio and Schanzenbach

[12] More details on our data sources are in Appendix B.

[13] For more information, see https://nces.ed.gov/nationsreportcard/.

[14] We have also estimated results for 8th grade, and generally get results consistent with those from 4th grade. Because 4th grade outcomes allow for additional observations (due to the lesser time lag between pre-K and 4th grade), the 4th grade estimates tend to be more precise than the 8th grade estimates. In addition, the 4th grade results are less likely to be affected by potential bias from district in-migration and out-migration of former pre-K students than the 8th grade results. As discussed in the appendix, the potential bias from migration appears to be modest at the 4th grade level.

2013; Grissmer, Flanagan, Kawat, and Williamson 2000; Rosinsky 2014). We employ, however, the restricted-access microdata, available to qualified researchers via license with the Institute of Education Sciences of the Department of Education (U.S. Department of Education 2015a,b). These microdata not only contain a wealth of information about individual students taking the NAEP and characteristics of the schools they attend, they also contain school and district identifiers that allow the data to be matched longitudinally over time and to be linked to external sources, such as the Department of Education's near-census of public schools, the Common Core of Data.[15]

The NAEP data provide our main outcomes of interest: math test scores, reading test scores, assignment to special education (i.e., the student has an Individual Education Plan), and a measure of whether children are over-age for their grade (a measure of grade retention). NAEP test scores are provided (and reported publicly) as a scale score; for the results we report in this paper, we convert the scale score to a percentile score using the 2013 NAEP score distributions for each grade and subject.[16] We implement this conversion because research by Chetty et al. (2011) has shown that percentile test scores are linearly related to adult earnings measured in dollars.[17] In addition, because previous research has found that pre-K programs may improve later life outcomes through their effect on socioemotional as well as academic skills (Heckman, Pinto, and Savelyev 2013), we also examine the assignment to special education and whether a

[15] To our knowledge, Fitzpatrick (2008) is the only previous paper to use the NAEP microdata to examine the effect of pre-K. However, she focuses on the implementation of Georgia's universal pre-K program and did not exploit within-state variation. Chingos (2015) demonstrates how the microdata can be used for a much richer set of controls to more accurately measure comparisons in performance across students.

[16] To minimize burden, individual students take only a portion of the full test, and item response theory is used to statistically impute multiple plausible scale scores for each student. We follow the literature and average these plausible scale scores for each student. The scale scores are approximately normally distributed.

[17] In practice, we obtain very similar results, quantitatively, if we use scale scores instead of percentiles, an apparent artifact of the NAEP scaling.

student is above the normal age cutoff for his or her grade.[18] These latter outcomes are more likely to capture learning difficulties that reflect nonacademic as well as academic deficiencies.

Common Core of Data

The CCD annually provides detailed characteristics of individual schools and school districts (local education agencies), including enrollment by sex, grade, and ethnicity, the share of students eligible for free and reduced-price lunch,[19] pupil-to-teacher ratios, type of locale, and others.[20] Of greatest utility for this paper, the CCD reports counts of pre-K enrollment within the public schools. This measure is not ideal, as it does not capture pre-K programs that are publicly funded but operate in centers outside the public schools. This measure also does not account for enrollment in private pre-K programs, which are in a few cases publicly subsidized (Barnett and Hustedt 2011).[21]

Nonetheless, we believe that enrollments from the CCD offer the best measure of spatial and temporal variation in the diffusion of pre-K. Some evidence suggests that pre-K programs located in public schools may be of higher average quality and lead to better results (Magnuson, Ruhm, and Waldfogel 2007), possibly because of better funding, better coordination with school expectations, and fewer transitions for children. Additionally, whereas previous papers (Cascio and Schanzenbach 2013; Fitzpatrick 2008) focus on the rollout of a universal pre-K program in one or two states, essentially making the adoption of pre-K into a binary event, the CCD counts

[18] More specifically, using the exact birthdate in the NAEP microdata, we define a student to be over-age if she turns 10 *prior* to the July 1st that immediately precedes the start of her 4th grade school year.

[19] The National School Lunch Program provides subsidized school lunches for students in families whose income falls below 185 percent of the federal poverty guidelines.

[20] Some of these characteristics are also reported in the NAEP itself, but they are missing for a nontrivial number of schools and districts. The CCD also allows district financial data, including spending per-pupil, to be matched to NAEP.

[21] Head Start, a federal preschool program intended for low-income students, may operate in partnership with public and private schools as well as standalone centers. We do not attempt to disentangle the source of funds used to pay for pre-K in the CCD enrollments.

offer changes in the intensive margin of pre-K for 50 states and the District of Columbia. This alone would provide advantages in estimation relative to previous studies, which typically employ few effective treatment groups and thus can suffer problems of inference (Conley and Taber 2011; Donald and Lang 2007). Furthermore, the CCD allows us to examine pre-K enrollment *at the district level*, as district codes can be matched to identifiers within the NAEP data set, something that has not been possible in previous research.[22] Shifting the unit of analysis from the state level to the district level enormously increases the size and precision of our "natural experiment." The additional variation from district data permits detection of even modest pre-K effects. As pre-K needs only modest effects to pass an expected-benefit–cost test, the increased precision of our estimates should be invaluable to policymakers.

Quality Indicator

To deal with the difficult issue of capturing pre-K "quality," we rely on the opinions of outside experts. Specifically, we draw upon a report by the Gates Foundation, which identified four exemplary programs: those in Boston, Maryland, New Jersey, and North Carolina (Minervino 2014). Other research has also identified New Jersey (Barnett et al. 2013), Boston (Weiland and Yoshikawa 2013), and North Carolina (Ladd, Muschkin, and Dodge 2014) as effective programs. Therefore, we classify New Jersey, Massachusetts, Maryland, and North Carolina as high-quality states.

In addition, other research has directly shown that Tulsa has a high-quality pre-K program (Phillips et al. 2009). Oklahoma's overall state pre-K program encourages many of the

[22] As a check on our district-level results, we also estimate similar equations at the school level, which examines how pre-K enrollment at a specific school affects 4th grade outcomes at the same school five years later. Because of student mobility across schools, and because some districts may concentrate their pre-K programs in selected schools, we regard these school results as less reliable. Nonetheless, they are generally consistent with our district results.

same features that relate to Tulsa's quality, such as requiring early education teaching credentials and paying pre-K teachers the same wages as other public school teachers. Therefore, we also treat Oklahoma as a state with "quality" pre-K programs, giving us a total of five "quality" states.

We acknowledge that such a binary quality indicator is somewhat arbitrary, and potentially subject to manipulation by the researcher. For example, an unscrupulous researcher could experiment with many possible "quality state" groups and report results only for the quality state group that gave the desired empirical results. This would both bias the coefficient estimates and invalidate the reported standard errors. To avoid this problem, we specified our quality state indicator before any empirical examination of a specific state's programs, and it was the only quality state grouping we considered. This prespecified design should make it more likely that our results are representative and valid.

These indicators of "quality" states are used in our later specifications to see whether the expansion of district pre-K in "quality states" has a greater effect in improving 4th grade outcomes compared to pre-K expansion in other states. More details on our empirical specifications are provided below.

Comparing Pre-K Data Sources

Because our choice of pre-K enrollment is uncommon (but not unprecedented) in the literature, we have examined how the CCD measure compares to two more widely used pre-K enrollment measures: the state-level counts tabulated by NIEER (various years) and enrollment rates derived from the census and the American Community Survey.

To convert the counts provided in the first two sources into a rate comparable with the third source, we construct either *population shares* or *population ratios*, depending on the level of analysis. At the state level, we divide the annual count of students enrolled in pre-K programs in public schools by the annual estimate of a state's four-year-olds, as provided by the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute.[23] We do this for counts from both the CCD and NIEER. Thus, these population shares represent the fraction of a state's four-year-olds enrolled in a public pre-K program in a given year. At the district level, however, there is no reliable and consistent source for the annual count of four-year-olds. We therefore construct a population ratio with the CCD data by dividing the count of pre-K enrollment by the count of 1st grade enrollment at the same district in that year.[24] These population ratios by district can be aggregated to the state level, weighting by 1st grade enrollment. (We use similar calculations at the school level for some sensitivity analyses.)

Table 1 shows how these measures correlate at the state-year level. Not surprisingly, the CCD state population shares (1) and population ratios at the levels of state (2), district aggregated to state (3), and school aggregated to state (4) all correlate very highly, with $r > 0.95$. But each of the CCD measures in turn also correlates highly with the NIEER state-funded pre-K rate, with $r > 0.75$. The CCD measures also correlate strongly with the ACS public enrollment rate of four-year-olds, with $r > 0.55$. Reassuringly, the CCD measures do *not* significantly correlate with NIEER's enrollment statistics for Head Start, most of which takes place outside public schools.[25] The CCD pre-K enrollments thus appear to have ample external validity.

---

[23] The Surveillance, Epidemiology, and End Results (SEER) program, http://seer.cancer.gov/, processes population data from the U.S. Census Bureau to be used in calculating rates of cancer incidence in the population at the state and county levels. It produces a more consistent population series over time than the census estimates.

[24] Grissmer et al. (2000) employ this technique at the state level. At smaller geographies, there is a chance that this ratio exceeds unity, but empirically this occurred only in about 3 percent of cases. Functionally, we recoded ratios above 1 but less than 1.5 to unity, and we dropped observations with ratios of 1.5 or greater, although the results are not sensitive to these restrictions.

[25] The correlation between the census/ACS measure and NIEER's Head Start statistic is higher, which is also plausible, as many families filling out the census/ACS may consider Head Start as public school enrollment.

Table 1 Correlations of Pre-K Measures across Data Sources, at State-Year Level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) CCD state PK share of 4-year-olds	**1.000**						
(2) CCD state PK/G1 ratio	0.976	**1.000**					
(3) CCD district PK/G1 ratios (aggregated)	0.957	0.981	**1.000**				
(4) CCD school PK/G1 ratios (aggregated)	0.949	0.962	0.980	**1.000**			
(5) NIEER state PK share of 4-year-olds	0.768	0.752	0.759	0.821	**1.000**		
(6) NIEER Head Start share of 4-year-olds	0.098	0.060	0.052	0.090	0.195	**1.000**	
(7) Census/ACS share of 4-year-olds	0.559	0.582	0.590	0.592	0.600	0.396	**1.000**

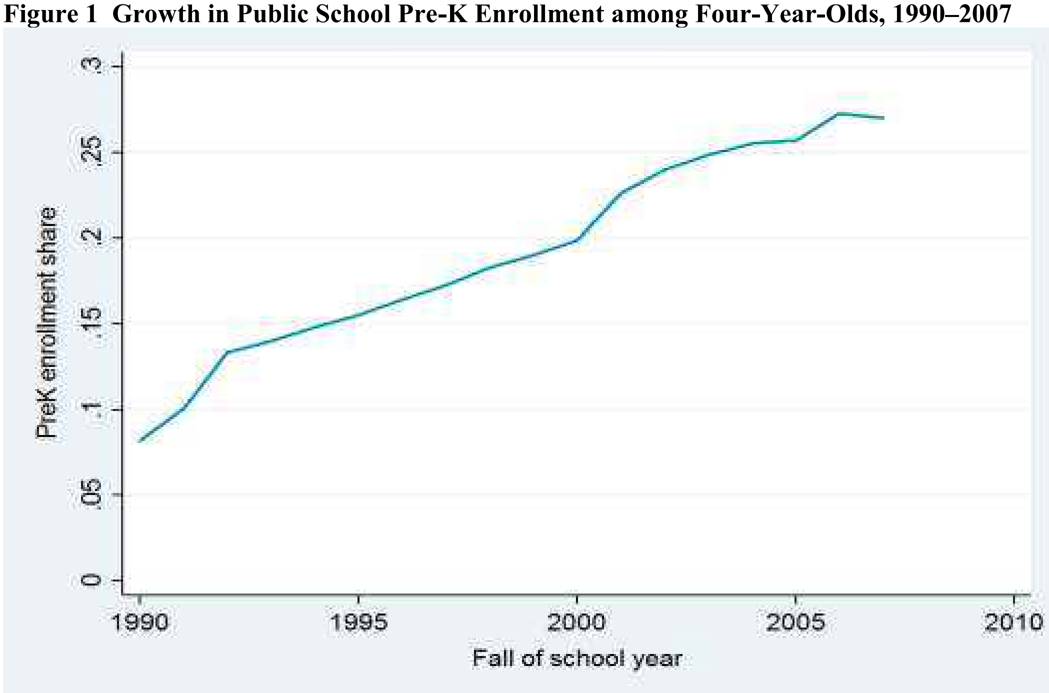
SOURCE: Authors' calculations from the Common Core of Data (various years), NIEER State Preschool Yearbooks (various years), 1990 and 2000 census and American Community Surveys (various years).
NOTE: Pairwise Pearson correlations are calculated at the state-year level for all valid state-year pairs. CCD data cover fall 1990 through fall 2007 school years, NIEER data cover fall 2001 through fall 2007 school years, and census/ACS data cover spring 1990 (matched to fall 1990 in CCD), spring 2000 (matched to fall 1999 in CCD), and fall 2001 through fall 2007. The ACS enrollment share matched to the fall of each year t is a weighted average of the ACS fielded in year t (0.375) and year $t+1$ (0.625) to approximate coverage for the school year. CCD ratios are calculated by summing the numerator within unit, summing the denominator within unit, taking the quotient, and then averaging using the denominator as weights. We do not use data beyond the fall of 2007, as that is the latest year that can be matched to 4th grade outcomes in NAEP.

Considering Pre-K Time Trends at Different Geographies in the CCD

Figure 1 shows the CCD pre-K population share for the United States between 1990 and 2007. (This time frame corresponds to our 4th grade NAEP sample measured five years later.) In the fall of 1990, approximately 8 percent of four-year-olds were enrolled in pre-K at a public school. This share steadily rose over time, reaching 27 percent by 2007, and is consistent with aggregate patterns documented by NIEER.

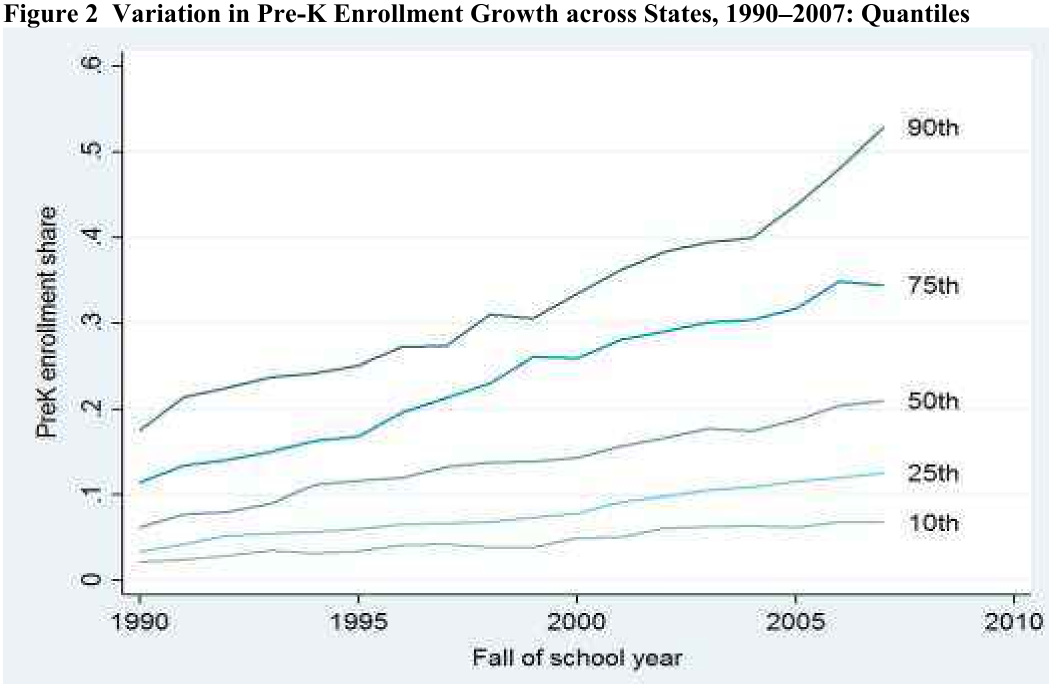
This increase, however, has not been equally distributed across states or districts. Figure 2, for example, plots pre-K enrollment shares by quantiles of states. The 10th percentile state (or the state with the fifth-lowest share of pre-K enrollment in a given year) grew its enrollment share from 2 percent in 1990 to about 8 percent in 2007. In contrast, the 90th percentile state increased its enrollment share from 18 percent to over 50 percent in the same period. Thus, while pre-K enrollment was increasing broadly over time, it increased faster in some states than others, leading to greater dispersion.

Figure 1 Growth in Public School Pre-K Enrollment among Four-Year-Olds, 1990–2007



NOTE: Figure shows national counts of public school pre-K enrollment (from the CCD) normalized by the number of four-year-olds (from SEER data).

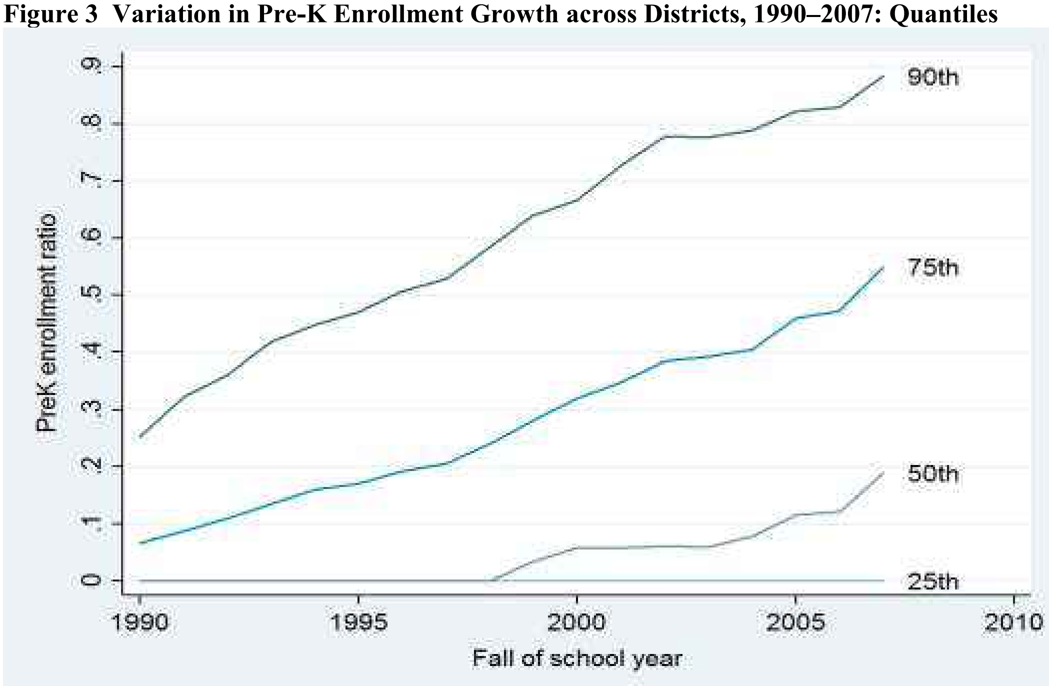
Figure 2 Variation in Pre-K Enrollment Growth across States, 1990–2007: Quantiles



NOTE: Figure shows specific quantiles among U.S. states in public school pre-K enrollment shares (normalized by the population of four-year-olds). For example, the 90th percentile shows the enrollment share for the state with the fifth-highest pre-K share each year. For sources, see Figure 1.

A similar but even more dramatic pattern exists for school districts, as shown in Figure 3. Now using enrollment ratios rather than population shares, the bottom quarter of districts have no public pre-K over the entire time horizon, and even the median district did not begin offering pre-K until the late 1990s. On the other hand, enrollment in the 75th percentile district jumped from 8 percent to 55 percent, and the 90th percentile district shot up from 26 percent to nearly 90 percent. Thus, dispersion in pre-K enrollment has grown much faster across districts than across states. For estimation, these disparate changes in district pre-K enrollment provide a natural experiment that helps us estimate effects on 4th grade student outcomes. Even after we control for district and year fixed effects, there is considerable variation in district pre-K enrollment ratios, allowing for more precise estimation than has been possible in previous studies.

Figure 3 Variation in Pre-K Enrollment Growth across Districts, 1990–2007: Quantiles



NOTE: Figure shows specific quantiles among U.S. public school districts in public school pre-K enrollment shares (normalized by the district's grade 1 enrollment), using CCD data. For example, the 90th percentile shows the minimum enrollment share for the top tenth of districts each year.

Analytic Samples

We construct our analytic samples by merging the pre-K enrollment measures from the CCD with NAEP data. Because students taking the 4th grade NAEP would have been enrolled in pre-K five years earlier, assuming normal grade progression, our matching procedure incorporates this lag. Given the NAEP administrations for each state and subject and the availability of pre-K enrollment from the CCD, Appendix Tables B1 and B2 show valid state-year combinations that compose the analytic samples.[26]

Because the NAEP data are at the student level and the CCD pre-K data—which provide the source of identifying variation—are at the district level (or for some estimates, the state or school level), we collapse the NAEP data to cells defined by NAEP test year, grade, test subject (math or reading), and geographic unit. We describe the details of this step in the empirical strategy section, below.[27] This produces samples at the district-year level (or state-year level and the school-year level in some specifications). While the NAEP data can be matched to CCD data for all test years at the state level, the matching at substate levels relies on the district and school identifiers in the restricted NAEP, which are missing in a few instances.[28] On average, the data for a given state, test year, grade, and subject are based on approximately 70 districts, about 140 schools, and about 3,100 students. Similarly, a given district-year cell generally consists of at least 30 students, and sometimes—for large districts—many more.

[26] A few states (and their constituent districts and schools) do not report pre-K enrollment in some years, which is the source of the blanks from 2003 onward. Notably, California never reports pre-K enrollment at the district or school level, or by race at the state level.

[27] Fitzpatrick (2008), the other study of which we are aware that uses restricted NAEP data, estimates the effect of pre-K directly using the NAEP microdata.

[28] We successfully matched 100 percent of districts identified in the NAEP to the CCD, but because some schools in the NAEP lacked the school identifiers used in the CCD, we could match only 94 percent of NAEP schools (across all years) to the CCD.

Methodology

Our augmented differences-in-differences strategy employs a two-stage design to estimate the effects of pre-K access on students' academic and behavioral outcomes. The first-stage uses the NAEP microdata to regress student-level outcomes on student-level covariates and a vector of geography-year indicator variables. The coefficients on these dummies, which represent means of the outcome variable adjusted for student characteristics, become the outcome variables for the second stage. The second stage, in turn, regresses these adjusted means on the appropriate pre-K measure and other covariates to identify the causal impact of pre-K. Donald and Lang (2007) demonstrate that such a two-stage approach can yield better inference when the number of groups is small; it also is computationally simpler.

More specifically we first estimate the equation:

$$y_{ig} = \boldsymbol{X_{ig}}\boldsymbol{\alpha} + \boldsymbol{Z_g}\boldsymbol{\gamma} + \varepsilon_{ig}\,, \tag{1}$$

where y_{ig} is a student-level test score in math or reading, an indicator variable for whether the student receives special education services, or an indicator for above the normal age cutoff for 4th grade, with i indexing students and g indexing geography (state, district, or school). $\boldsymbol{X_{ig}}$ is a vector of student characteristics, including binary indicators for sex, race, participation in the federal assisted lunch program (separately for free and reduced-price), and whether the student is an English-language learner. $\boldsymbol{Z_g}$ is a vector of indicator variables for geography. Finally, ε_{ig} is a student-level error term. Equation (1) is estimated separately for each NAEP year and outcome variable, allowing the relationship between student characteristics and outcomes to vary over time and across outcome variables. Our main results reported in the text are for 4th grade outcomes.

The coefficient estimates $\boldsymbol{\gamma}$, which we stack across years for each outcome variable, are geography-specific fixed effects, net of student characteristics. We reparameterize this vector (within year and outcome variable) by subtracting the overall weighted mean outcome for the entire sample so that the new vector represents deviations from the national mean (and thus sums to zero).

The reparameterized vector $\tilde{\boldsymbol{\gamma}}$ becomes the outcome variable in the second stage:

$$\tilde{\gamma}_{gt} = \beta_0 + \beta_1 \cdot PreK_{g,t-\ell} + \boldsymbol{\mu_g} + \boldsymbol{\varphi_t} + \boldsymbol{Q_{gt}\theta} + \nu_{gt}\,, \tag{2}$$

where $PreK_{g,t-\ell}$ is the measure of pre-K in geography g, lagged the appropriate number of years to correspond to when the test cohort would have been enrolled in pre-K, $\boldsymbol{\mu_g}$ is a vector of geography dummies, $\boldsymbol{\varphi_t}$ is a vector of test year dummies, $\boldsymbol{Q_{gt}}$ is a vector of time-varying characteristics of the geography, and ν_{gt} is an idiosyncratic error term, which we allow to be arbitrarily correlated within geography. Equation (2) is estimated separately by outcome. The coefficient of interest is β_1, which shows how the normalized outcome changes when the pre-K measure varies from 0 (no pre-K) to 1 (presumed to be full, or universal, pre-K).

In the discussion of our results, we emphasize our implementation of this two-step procedure at the level of districts, as this gives us the most precision in estimation but avoids problems in student mobility across schools. But for comparison purposes, we also in the text report some results at the state level (broken down either by student race or student eligibility for the assisted lunch program).[29] This type of state analysis is akin to the studies by Cascio and Schanzenbach (2013) and Rosinsky (2014), except we use the CCD to measure pre-K intensity.

[29] For the race analysis, we use race-specific pre-K enrollment, as this is reliably available at the state level.

Table 2 presents descriptive statistics for the main variables of interest used in our state analyses. These include the main independent variable of interest, the pre-K enrollment rate or share, and the various dependent variables.

Table 2 Summary Statistics for State-Level FRL Samples, Grade 4

Variable	Low-income students		Non-low-income students	
	Mean	SD	Mean	SD
Pre-K share	0.187	0.145	0.220	0.175
Math percentile score, raw	34.3	7.2	55.4	7.7
Math percentile score, student-adjusted	−7.9	4.6	4.9	4.1
Reading percentile score, raw	36.0	4.9	57.6	4.6
Reading percentile score, student-adjusted	−7.9	3.7	5.4	3.7
Special education share (×100), raw	16.7	3.8	11.0	2.3
Special education share (×100), student-adjusted	3.8	3.6	−2.7	2.0
Over-age for grade (×100), raw	13.8	6.4	6.0	3.0
Over-age for grade (×100), student-adjusted	3.6	5.9	−3.1	2.8

NOTE: All statistics are weighted by the number of NAEP students contributing to the relevant cell; unweighted statistics are similar. "Raw" statistics shown are as calculated in the NAEP data; adjusted statistics (used in the analyses) are recentered to have a weighted mean of 0 in each test year; see text for details. Low-income students are those eligible for free or reduced-price lunch, as indicated in the NAEP data, while non-low-income students are those ineligible for the lunch program. Sample sizes in the second stage (at the state-year level) are approximately 380 for math, 390 for reading, 470 for special education, and 430 for over-age. Cell sizes in the first stage—the number of students contributing to the outcome mean at the state-year level for each income group—average about 1,640 for non-low-income students (min=40, max=3,990 across state-years), and 1,580 for low-income students (min=20, max=7,060). All sample sizes are rounded to the nearest 10 to accord with disclosure restrictions.

Analysis at the district level provides significantly greater variation in pre-K than is possible with a state-level design. Importantly, it also allows for greater control of possible unobservables that can bias estimates. Because many districts are sampled multiple times across NAEP test years, we can include district fixed effects to net out permanent differences across these geographies.[30] Moreover, the CCD also permits us to control for time-varying characteristics of districts, including the share of students eligible for the assisted lunch program (categorical), racial and ethnic composition (categorical), and whether the district (school) is in

[30] Due to the large number of districts, we use the -reghdfe- package in Stata (Correia 2014) to implement the fixed effects.

an urban area, suburban area, town, or rural area. Table 3 presents select descriptive statistics for our district-level sample (see also Figure 3).

Table 3 Summary Statistics for District-Level Samples, Grade 4

Variable	All districts		Majority-black districts		Majority-Hispanic districts		90% + white districts	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Pre-K share	0.206	0.230	0.350	0.293	0.389	0.287	0.151	0.216
Share in "quality" states	0.118	0.323	0.115	0.320	0.039	0.193	0.057	0.232
Math percentile score, raw	46.3	13.1	29.9	9.5	39.3	9.3	53.1	11.3
Math percentile score, adjusted	−0.6	7.6	−0.8	5.5	3.7	7.7	−2.9	8.1
Reading percentile score, raw	47.3	11.9	32.9	8.7	38.6	8.3	54.4	10.0
Reading percentile score, adjusted	−0.5	6.8	−0.8	5.3	2.6	6.3	−2.3	7.7
Special educ. share (×100), raw	13.9	6.1	13.5	5.8	12.6	5.0	14.7	7.5
Special educ. share, adjusted	0.7	5.7	−2.1	5.5	−1.3	4.9	2.1	6.8
Over-age for grade, raw	9.9	7.4	15.9	7.6	12.6	6.3	8.3	7.1
Over-age for grade, adjusted	0.4	6.5	2.3	6.8	0.5	6.6	0.7	6.6
District per-pupil spending (000s)	7.3	1.9	8.1	1.6	6.7	1.7	7.9	1.9
N (district-years)	23,450		1,400		920		9,610	
Unique districts	5,790		330		200		2,550	

NOTE: All statistics are weighted by the number of NAEP students contributing to the relevant cell; unweighted statistics are similar. "Raw" statistics shown are as calculated in the NAEP data; adjusted statistics (used in the analyses) are recentered to have a weighted mean of 0 in each test year; see text for details. "Quality" states include MD, MA, NJ, NC, and OK. Per-pupil district spending is taken from the Common Core of Data and adjusted for inflation (to $1,999) and the comparable wage index (across districts) via Taylor and Fowler (2006). The number of observations and unique districts vary slightly across outcomes; the statistics shown are the maximum across outcomes. Average cell sizes in the first stage—the number of students contributing to the outcome mean at the district-year level for each group—range from about 20 for rural districts and 90+ percent white districts to about 120 for city districts; the overall average is about 40. All sample sizes are rounded to the nearest 10 to accord with disclosure restrictions.

In addition, we include controls meant to capture the availability of alternative options to public pre-K and the cumulative district spending between pre-K and 4th grade. Specifically, we control for the number of Head Start slots for four-year-olds within 10 kilometers of any public school in the district and the number of private school preschool slots within 5 kilometers of any public school in the district; we normalize the count of slots in each case by the district's 1st grade enrollment.[31] Furthermore, we control for a quartic in district public school spending per

[31] The number of slots and location of Head Start programs is taken from the Head Start Program Information Reports database, available from the Office of Head Start within the U.S. Department of Health and

student averaged over the years from pre-K to 4th grade.[32] This allows us to control for K–12 spending that may be correlated with pre-K expansion and that could also affect 4th grade outcomes. For example, some districts' investments in pre-K may come at the expense of reduced K–12 spending; alternatively, districts investing in pre-K may be increasingly pro-education districts that are also raising K–12 spending.

As alluded to earlier, in some specifications we also interact the pre-K enrollment rate with an indicator for whether the school district is in one of our five prespecified "quality states." The estimation equation then can be written as follows:

$$\tilde{y}_{gt} = \beta_0 + \beta_1 \cdot PreK_{g,t-\ell} + \beta_2 \cdot PreK_{g,t-\ell} \cdot Quality_s + \boldsymbol{\mu_g} + \boldsymbol{\varphi_t} + \boldsymbol{Q_{gt}\theta} + \nu_{gt}\,, \tag{3}$$

Of interest here is whether district pre-K has greater effects in states that are specified to have quality programs (i.e., $\beta_2 > 0$), as we would expect. Also of interest is the net effect of pre-K at the district level in quality states ($\beta_1 + \beta_2$), and in the remaining states (β_1).

For the state-level regressions, our identifying variation comes from changes in pre-K enrollment within a state over time. For the district regressions, the identifying variation comes from within-district changes in pre-K enrollment over time.[33] In the latter cases, in addition to examining average effects of pre-K across districts, we also examine heterogeneity over certain types of districts. In the text, we report results when districts are classified by their racial

Human Services (2010). The number of slots and location of private preschool programs is taken from the Private School Universe Survey within the U.S. Department of Education (2015a,b). Distances between Head Start or private school programs and public schools are calculated from latitude and longitude after geocoding using the -geodist- package in Stata. Head Start unfortunately tracks the locations only of grantee agencies and not necessarily where services are actually delivered, so we increased the distance threshold from 5 to 10 kilometers for Head Start slots.

[32] Specifically, we use current operating expenditures divided by total (nonadult) enrollment from the CCD. We adjust the spending measures to account for geographic price variation and inflation using the comparable wage index maintained by Lori Taylor (see http://bush.tamu.edu/research/faculty/Taylor_CWI/).

[33] In some sensitivity tests, we add state-by-year dummies to the district regressions. In this case, identification implicitly comes from changes in district enrollment over time, relative to what is observed for other districts in the same state over the same time period. This comparison quite strains the data and tends to reduce precision.

composition. In the appendices, we report results when districts are classified by the share of students on free or reduced-price lunch, and by total enrollment.[34]

Although we consider our approach to have several advantages over earlier studies, it is not without a set of disadvantages. First, even though we can better account for possible endogeneity in the pre-K expansion, we cannot eliminate it entirely. If individual districts expand pre-K because test scores are trending downward, our methodological approach will not adequately control for it.[35] This could lead to the possibility of some downward bias to our estimated pre-K effects. We do not think this is likely to be a major problem, as bias would result only if the same time trends in prior years that caused districts to expand pre-K were persistent enough to cause cohort test score effects in the NAEP, five years later.[36] Since we always control for district fixed effects, we should avoid endogeneity biases due to persistent district characteristics being correlated with pre-K enrollment rates.

Second, due to the nature of our sample, there may be some attenuation in our estimates due to in-migration and out-migration of students between the pre-K year and the NAEP test year. This migration problem is likely to be highest at the school level, followed by the district level, with state-level analyses having the least migration issues. If migration is random, and if there are no spillover effects of one student's skills on other students, then our estimates will be biased toward zero, as the true effect of pre-K on individual test scores and other outcomes will be equal to our estimate divided by (1 − migration rate). In the appendix, we examine the migration problem at the school district level, using data from the Panel Study of

[34] In each case, we use categorical indicators based on sample averages of the characteristic. These factors are among those that have been identified in previous research as showing heterogeneous treatment effects (Cascio and Schanzenbach 2013; Fitzpatrick 2008).

[35] Since NAEP results are not released by district, this is problematic only to the extent that NAEP results correlate with other state and district exams, for which evidence is mixed (Reardon, Kalogrides, and Ho 2016).

[36] In the appendix we also conduct tests for whether past NAEP test scores predict current pre-K enrollment rates at the district level. Effects are generally statistically insignificant and small.

Income Dynamics (PSID), and conclude that district migration rates from pre-K to 4th grade are typically of modest size, approximately 20 percent.

Third, we do not explicitly account for the specific program quality of a district's pre-K programs, including length of school day, as there is no measure of quality available for every district. Our overall results (without the state quality indicator) will capture an average treatment effect of all public pre-K programs as they were implemented, and such an average treatment effect may mask strong positive impacts from some programs and negative impacts from others. Our quality state indicator detects whether this average treatment effect is stronger in states that are thought to have higher-quality pre-K programs. Ideally, we would have district-specific quality indicators.

RESULTS

State-Level Results by Student Income

Our main results focus on effects of district-level pre-K. But for comparison with prior studies, we also present results from estimating Equation (2) at the state level, separately by student income level. Table 4 reports these results. We examine four outcomes across rows: 1) the adjusted percentile math score, 2) the adjusted percentile reading score, 3) the adjusted share of students reporting an Individual Educational Program (i.e., receiving special education services), and 4) the adjusted share of students above the normal age cutoff for 4th grade. The columns show results differentiated by whether a student is eligible for a free or reduced-price lunch (e.g., below or above 185 percent of the federal poverty guideline).

When interpreting these results, we should keep in mind what the estimated effects would have to before pre-K to pass a benefit-cost test. As discussed above, at the average pre-K

Table 4 The Effects of Pre-K on State-Level 4th Grade Outcomes: FRL vs Non-FRL students

	(1)	(2)
	FRL students	Non-FRL students
Math scores (percentile)	3.513	2.910
	(2.100)	(2.684)
Reading scores (percentile)	−0.546	−1.612
	(2.282)	(2.946)
Special education (proportion ×100)	0.33	1.15
	(1.77)	(1.25)
Over-age for grade (proportion ×100)	−2.56	2.32
	(2.52)	(1.40)

NOTE: Each cell is from a separate regression of the NAEP outcome on the pre-K measure, a set of state dummies, a set of test year dummies, and a quartic in cost-of-living adjusted current spending per student, averaged over the test year and preceding four years to account for time since pre-K. Each observation is a state-year, and there are 381 for math, 393 for reading, 470 for special ed., and 430 for over-age for grade. Standard errors in parentheses are clustered by state. The underlying dependent variables are state-year cell means that have been regression-adjusted for individual student characteristics and recentered so that the national weighted mean is zero for each year; see text for details. Column (1) shows results for students who are eligible for free or reduced-price lunch (FRL; family income below 185% of poverty line) and column (2) shows results for ineligible students, where eligibility is taken from the NAEP student-level data. The independent variable is the ratio of pre-K enrollment in that state-year to first grade enrollment in that same state-year, taken from the Common Core of Data. The coefficients thus reflect the estimated effect of moving from 0 to 100 percent enrollment in pre-K.

program cost, this would require a positive 1.3 percentile impact on test scores, a 3.3 percentage point reduction for special education, and a 2.7 percentage-point reduction for over-age for grade.

None of the state-level estimates in Table 4 is close to being statistically significantly different from zero. Although the point estimates for math test scores are two to three times the benefit-cost threshold, they are imprecisely estimated, and the confidence interval implies the benefit-cost ratio could exceed six or fall below zero. The point estimate for over-age for grade for free- and reduced-price lunch students is slightly less negative than needed to just pass a benefit-cost test, but it is also imprecisely estimated. All other estimated effects are of the "wrong sign," but are again insignificantly different from zero. Indeed, the effects on reading are sufficiently imprecise that their passing a benefit-cost test cannot be ruled out despite the negative point estimate.

In examining these results, the most salient take-away is simply the imprecision of results, even though we are aggregating across numerous states and years. The standard errors are of the same order of magnitude as the cutoffs for benefits equaling costs. The implied confidence intervals are large enough that it is rarely possible to reject either zero effects or substantively meaningful effects. The underlying difficulty is that pre-K may have social benefits exceeding costs even if its effects are only modestly sized, and pooled data on states simply do not yield the needed precision.

In terms of magnitude, the point estimates are smaller than have been found in previous studies using state-level variation. On the one hand, we might expect larger point estimates, as many previous studies used a dichotomous indicator for pre-K while we use a continuous one, and even in states such as Georgia and Oklahoma that adopted large-scale public pre-K programs, participation among four-year-olds was far from universal. Thus, as a matter of scaling alone, the estimates we show should approximately be halved to be commensurate with those from many of the earlier studies. On the other hand, it is quite possible that pre-K exposure averaged across different quality programs yields smaller net effects. We note, however, that our estimates are of comparable precision (and well within the confidence intervals) of those in Fitzpatrick (2008) and Cascio and Schanzenbach (2013) once the Conley-Taber adjustments are applied. State-level estimates of the effects of average pre-K programs are simply not precise enough for policy purposes.

District-Level Results

We now turn to estimates using district-level variation in pre-K. Since approximately 70 districts are sampled from each state on average, the effective number of observations and identifying variation is much larger than in the state-level results.[37]

The average effects of pre-K across all districts are shown in Table 5, with rows delineating the four outcomes and columns delineating specifications with and without district time-varying controls. All estimates control for both year and district fixed effects.

Table 5 The Effects of Pre-K on District-Level 4th Grade Outcomes

	(1)	(2)
Math scores (percentile)	−0.114	0.168
	(0.672)	(0.646)
Reading scores (percentile)	−1.594**	−1.301**
	(0.625)	(0.580)
Special education (proportion ×100)	−0.78	−0.91
	(0.63)	(0.62)
Over-age for grade (proportion ×100)	−0.55	−0.49
	(0.47)	(0.44)
Include district fixed effects	Yes	Yes
Include district time-varying controls?	No	Yes

NOTE: * significant at the 0.10 level; ** significant at the 0.05 level; *** significant at the 0.01 level. Each cell is from a separate regression of the outcome on the pre-K measure, a set of test year dummies, a quartic in cost-of-living-adjusted, district-level current spending per student (averaged over the test year and preceding four years to account for time since pre-K), and the other controls as shown. District time-varying controls include: categorical dummies for the share of students eligible for free or reduced-price lunch, the student enrollment (size) of the district, the share of students who are black in the district, the share of students who are Hispanic in the district, the number of private school pre-K slots available within 5 km of any school in the district (normalized by the district's grade 1 enrollment), and the number of Head Start four-year old slots available within 10 km of any school in the district (normalized by the district's grade 1 enrollment). Each observation is a district-year, and there are 19,320 observations (5,280 unique districts) for math scores; 21,460 (5,520) for reading scores; 23,450 (5,790) for special education; and 23,330 (5,760) for over-age for grade. All observation and district counts have been rounded to the nearest 10 to comply with disclosure restrictions. Standard errors in parentheses are clustered by district. The underlying dependent variables are district-year cell means that have been regression-adjusted for individual student characteristics and recentered so that the national weighted mean is zero for each year; see text for details. The independent variable is the ratio of pre-K enrollment in that district-year to first grade enrollment in that same district-year, taken from the Common Core of Data. The coefficients thus reflect the estimated effect of moving from 0 to 100 percent enrollment in pre-K.

[37] Technically, schools are sampled, not districts. Thus, in several cases different schools within the same district are sampled over time. To the extent that there is significant variation in schools within a district (as is more likely with larger districts), estimation results may be confounded by compositional change. We attempt to address this issue by controlling for individual characteristics in the first stage. Using school-level results can obviate this issue, but at the significant cost of potential biases due to mobility across schools within districts, and due to some districts using limited number of schools as pre-K centers.

As with the state-level estimates in Table 4, the district-level estimates in Table 5 show little evidence of *statistically* significant benefits from the average pre-K program. However, with many more districts than states, the district-level estimates are much more precise than the state-level estimates: standard errors are less than one-third the magnitude of standard errors in the state-level estimates. As a result of the district estimates' greater precision, it is possible to more definitively rule out *substantively* "large" benefits. For example, the effects on special education and over-age for grade are precisely enough estimated that the confidence interval does not come close to the cutoff for benefits exceeding costs. Reading test scores are statistically significantly less than zero, so clearly the confidence interval excludes a substantively large positive benefit. For math test scores, only the estimate with district-varying controls could possibly indicate a substantively large benefit, and only just barely ($0.168 + 1.96 \times 0.646 = 1.43$, just larger than the benefit-cost threshold of 1.4 percentiles). But even here, we can rule out math test score effects that would yield a benefit-cost ratio of 2 to 1, let alone the very high benefit-cost ratios of 8 to 1 (or more) estimated for earlier pre-K programs such as Perry and the Chicago Child Parent Center.

What about effects of districts expanding pre-K in states that are believed to have high-quality programs? These results are shown in Table 6 and provide some evidence that quality matters.

In particular, the district expansion of pre-K in quality states raises 4th grade math scores by an amount that is both statistically significant and substantively important. The point estimate implies that for school districts in quality states, shifting from no pre-K to 100 percent coverage will increase math test scores by 2.8 percentiles. This is twice the cutoff for pre-K to have

Table 6 The Effects of Pre-K on District-Level 4th Grade Outcomes, Quality States vs. Other States

	Other states	Quality states	Difference
Math scores (percentile)	−0.469	2.840**	3.309**
	(0.773)	(1.206)	(1.469)
Reading scores (percentile)	−1.815**	0.638	2.453
	(0.656)	(1.321)	(1.492)
Special education (proportion ×100)	−1.02	−0.49	0.53
	(0.73)	(1.01)	(1.23)
Over-age for grade (proportion ×100)	−0.17	−1.59	−1.42
	(0.48)	(1.05)	(1.14)
Include district fixed effects		Yes	
Include district time-varying controls?		Yes	

NOTE: * significant at the 0.10 level; ** significant at the 0.05 level; *** significant at the 0.01 level. Each row is from a separate regression of the outcome on the pre-K measure, a set of test year dummies, a quartic in cost-of-living-adjusted, district-level current spending per student (averaged over the test year and preceding four years to account for time since pre-K), and district time-varying controls (see note to Table 5). The coefficients across columns show the pre-K measure, its interaction with an indicator variable for being a "quality program state" (equal to 1 for MD, MA, NJ, NC, and OK), and the net effect of pre-K in quality states. The estimates in column (1) thus shows the impact of moving from 0 to 100 percent enrollment in pre-K in all but the "quality states;" the estimates in column (2) shows the impact of moving from 0 to 100 percent enrollment in pre-K in the "quality states;" and the estimates in column (3) show the difference in these impacts. Each observation is a district-year; for sample sizes, see note to Table 5.

expected earnings benefits greater than costs.[38] Adding the "quality-state interaction" also tends to increase the reading test scores benefits of pre-K, as well as increasing the reduction in grade retention due to pre-K, although the differences across school districts in "quality" and other states are not statistically significant.

Of course, heterogeneity in pre-K impacts is likely to extend beyond our admittedly crude (if *ex ante*) measure of program quality. In Table 7, we show results for three different types of school districts: majority-black districts, majority-Hispanic districts, and districts whose student body is at least 90 percent white.[39]

[38] Technically, the ratio is slightly less, as average costs in the quality states are slightly higher than average.
[39] The appendix reports other results that differentiate districts by district size and district percentage eligible for a free or reduced-price lunch.

Table 7 The Effects of Pre-K on District-Level 4th Grade Outcomes, Quality States vs. Other States: by District Racial Composition

		Quality Interaction		
	Pooled	Other states	Quality states	Difference
Panel A: Majority Black Districts				
Math scores (percentile)	5.885**	5.778*	6.646*	0.868
	(2.901)	(3.121)	(3.965)	(4.315)
Reading scores (percentile)	3.828**	3.419*	7.415*	3.996
	(1.905)	(1.964)	(4.049)	(3.920)
Special education (proportion ×100)	−2.22	−3.16	5.15	8.31**
	(1.99)	(1.92)	(4.12)	(4.14)
Over-age for grade (proportion ×100)	−1.32	−1.90*	4.01	5.91
	(1.26)	(1.11)	(4.82)	(4.58)
Panel B: Majority Hispanic Districts				
Math scores (percentile)	−2.324	−3.699**	3.108	6.807**
	(1.913)	(1.850)	(3.202)	(3.421)
Reading scores (percentile)	−4.085	−5.337**	1.576	6.913
	(2.508)	(2.567)	(4.748)	(5.276)
Special education (proportion ×100)	−3.89*	−4.92**	−0.53	5.45*
	(2.00)	(2.08)	(2.61)	(3.07)
Over-age for grade (proportion ×100)	−0.78	−2.00	3.79	5.79
	(3.53)	(0.48)	(6.94)	(7.54)
Panel C: 90%+ White Districts				
Math scores (percentile)	−1.727*	−1.770*	0.086	1.856
	(0.899)	(0.910)	(4.288)	(4.340)
Reading scores (percentile)	−0.999	−1.251*	8.061**	9.312**
	(0.742)	(0.749)	(4.110)	(4.156)
Special education (proportion ×100)	−0.61	−0.71	3.37	4.08
	(0.65)	(0.66)	(3.16)	(3.20)
Over-age for grade (proportion ×100)	0.81*	0.75	2.07	1.32
	(0.46)	(0.47)	(1.49)	(1.54)

NOTE: * significant at the 0.10 level; ** significant at the 0.05 level; *** significant at the 0.01 level. See notes to Tables 4 and 5. Each panel represents regressions on the indicated subset of the data. All regressions include district fixed effects and district time-varying controls. The number of district-year observations—all rounded to the nearest 10 to comply with disclosure restrictions—for panel A is math, 1,130; reading, 1,270; special ed., 1,400; over-age, 1,340. The corresponding sample sizes for majority Hispanic districts are 760, 840, 920, and 910. For 90% plus white districts: 8,000, 8,800, 9,610, and 9,010. The number of unique districts is approximately between one-fourth and one-third the number of observations.

These regressions all control for fixed year and district effects, and also for time-varying district controls. For each racial category of districts, results are reported for each of the four outcomes from two different regressions. The leftmost column of numbers reports results when pre-K effects are restricted to be constant across all states (analogous to the estimates in Table 5).

The three columns to the right report results from specifications in which pre-K effects are allowed to differ between districts in quality states and other states (analogous to the estimates in Table 6).

For the estimates that impose constant pre-K effects across states, the most noteworthy result is for majority-black school districts: pre-K appears to increase math and reading test scores by an amount that is statistically significant and substantively important. A majority-black district shifting from zero pre-K enrollment to full pre-K improves its math scores by almost 6 percentiles and its reading scores by almost 4 percentiles. For these same majority-black districts, the point estimates suggest some possibility of pre-K reducing special education and grade retention by amounts that might be substantively important, but the estimates are too imprecise to distinguish the effects from zero.

In contrast, for heavily white districts, the pooled estimates in the leftmost column do not provide much evidence for benefits of pre-K. There is a modest negative impact on math test scores and a slight increase in grade retention, although these estimates are only statistically significantly different from zero at the 10 percent level. Even so, the precision is sufficient to rule out pre-K effects that would pass a benefit-cost test for any of the outcomes.

For majority-Hispanic districts, the pooled estimates are mixed. There is a weakly significant (albeit substantively large) reduction in special education assignments, with the point estimate implying full adoption of pre-K will reduce the share of students with individual education plans by almost 4 percentage points. For the other outcomes, however, the estimates are relatively noisy, wrong-signed for test scores, and not particularly informative.

For the results that allow district pre-K effects to vary across districts in quality states and other states, the point estimates often indicate greater pre-K effects on test scores in districts in quality states. In some cases, the estimated differences are statistically significant and/or substantively important. For heavily white districts, for example, full adoption of pre-K is estimated to increase reading scores by 8 percentiles in quality states, on par with the estimate for majority-black districts. For heavily white districts in other states, however, pre-K appears to reduce reading test scores at 4th grade. Expansion in quality states also increases reading scores in majority-black districts (7.4 percentiles, relative to 3.4 percentiles in other states). Interestingly, while pre-K expansion in majority-Hispanic districts in most states seems to have a statistically significant detrimental effect on math (−3.7 percentiles) and reading (−5.3 percentiles), the effects are positive, although noisy, in such districts in quality states.[40]

Overall, these district pre-K estimates are consistent with a reasonable story. Pre-K in the average state and for the average student and school is of insufficiently high quality to create large positive benefits that can be statistically detected. However, pre-K is of sufficiently high quality in the average state to create benefits for some disadvantaged students—notably, for students in majority-black school districts. Furthermore, in some high-quality states, pre-K can create benefits for broader groups of students.

The magnitude of some of these positive test score benefits for majority-black districts is roughly consistent with past meta-analyses, which have found average effects of about 5 percentiles around 4th grade (Camilli et al. 2010; Duncan and Magnuson 2013). However, many of the studies included in these meta-analyses are for smaller-scale programs targeted at disadvantaged students. The current study adds to this literature by suggesting that similar effects

[40] Pre-K expansion in quality states also seems to increase grade retention and special education assignments across the district types, although never statistically significantly so. This seems contrary to the test score results and may represent a greater likelihood of diagnosis through greater student monitoring, although given the imprecision of these interactions, we are hesitant to read too much into these patterns.

can sometimes be achieved for larger-scale public programs, throughout the nation, that serve disadvantaged populations.

DISCUSSION AND CONCLUSIONS

In this paper, we have used several data sets that together allow us to investigate the relationship between pre-K diffusion and educational outcomes on a representative sample of school districts throughout the country. Unlike most prior research, we do not examine the effects from a particular pre-K program or even a particular state's pre-K program; rather, we estimate the effects of all public pre-K programs averaged together, either for the whole country, or for groups of states that vary in expert opinion of the quality of their pre-K programs. The approach we use has advantages over previous geographic studies in providing far more identifying variation, controlling for more covariates that were potentially unobserved confounders, and producing national-level estimates. This last advantage also extends to randomized control trials of pre-K, which typically yield concerns of external validity over whether they generalize to other settings and time periods. On the other hand, our approach also has disadvantages relative to earlier studies. We do not directly observe individual-level treatment or short-term outcomes, as in control trial studies. And relative to both the control trial studies and the geographic studies, our measure of treatment is diffused because we pool many different programs together. Even for our results for districts in "quality states," there is presumably important variation in the quality of pre-K programs across different districts, and our estimation procedure does not capture this district-level variation. Put differently, whereas many prior studies looking at intensive or widely regarded programs analyzed what a pre-K program *could do* under the right circumstances, in this paper we effectively look at what typical pre-K programs *have done* in practice over the last two decades, both overall and in states with more highly regarded programs.

Our results indicate that pre-K programs in the public schools have done relatively little for the average student, school district, and state. However, pre-K programs do appear to have substantively large benefits when they are either higher quality or operated in more disadvantaged school districts, such as majority-black school districts. For these latter districts, some of the estimated positive effects of pre-K on test scores are equal to or greater than the effects suggested by meta-analyses of past small-scale programs.

Thus, we interpret the results as suggesting that large-scale pre-K programs *can* produce significant medium-term benefits. But both quality and context matter. Medium-term benefits are more likely if the program is high quality. Medium-term benefits are also more likely if the pre-K program operates in a context where students are more disadvantaged.

We do not view our results as being in contradiction with the positive impacts found in several earlier studies. As noted, many of the previous pre-K studies concentrated on specific programs that were likely of higher-than-typical quality, as suggested by both expert opinion and the magnitude of expenditures, and were also in many cases targeted at disadvantaged populations. Our results are also consistent with Rosinsky (2014), the only other study to our knowledge that looks at pre-K programs throughout the entire country, and that finds few positive benefits of the average state pre-K program.

However, we significantly add to the literature by finding that large-scale programs throughout the country can make a difference—with the right quality and context. Because much of the current policy debate is about the desirability of large-scale expansion of pre-K, these findings are highly policy relevant.

Our results are limited to medium-term outcomes, which may not always be predictive of long-term outcomes. From prior studies, even if the average pre-K program produces no measurable impact on 4th grade test scores, whether due to varied quality, test score fade-out, or both, it does not necessarily follow that there are no long-term or "sleeper" effects. As Heckman has noted on multiple occasions, pre-K may boost long-term social outcomes as much (if not more) through its effect on socioemotional skills as on academic ones. If these soft skills are not adequately captured in our NAEP proxies of special education and over-age for grade, future educational attainment and future earnings might be more greatly affected than predicted based on the medium-term results in the current paper. Therefore, for future research, we plan to explore the impacts of district pre-K on longer-term outcomes, such as high school graduation.

In addition, researchers should continue to seek better measures of pre-K quality that are more consistently correlated with outcomes and that can readily be used across studies and in policy work. Saying that "quality matters" for pre-K is a safe policy recommendation, but without highly predictive specific design features it does not provide much help to real-world policymakers in describing how to structure an effective pre-K program.

REFERENCES

Barnett, W. Steven, Megan E. Carolan, Jen Fitzgerald, and James H. Squires. 2014. *The State of Preschool 2014: State Preschool Yearbook*. New Brunswick, NJ: Rutgers University, National Institute for Early Education Research.

Barnett, W. Steven, Allison H. Friedman-Krauss,, G. G. Weisenfeld, Michelle Horowitz, Richard Kasmin, and James H. Squires. 2017. *The State of Preschool 2016: State Preschool Yearbook*. New Brunswick, NJ: Rutgers University, National Institute for Early Education Research.

Barnett, W. Steven, and Jason T. Hustedt. 2011. *Improving Public Financing for Early Learning Programs.* NIEER Policy Brief 23.

Barnett, W. Steven, Kwanghee Jung, Min-Jong Youn, and Ellen C. Frede. 2013. *Abbott Preschool Program Longitudinal Effects Study: Fifth Grade Follow-Up*. New Brunswick, NJ: Rutgers University, National Institute for Early Education Research.

Bartik, Timothy J. 2011. *Investing in Kids: Early Childhood Programs and Local Economic Development*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

———. 2014. *From Preschool to Prosperity: The Economic Payoff to Early Childhood Education*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Bartik, Timothy J., Jonathan Belford, William T. Gormley, and Sara Anderson. 2016. "A Benefit-Cost Analysis of the Tulsa Universal Pre-K Program." Mimeo.

Burchinal, M., K. Kainz, K., and Y. Cai. 2011. "How Well Do Our Measures of Quality Predict Child Outcomes? A Meta-Analysis and Coordinated Analysis of Data from Largescale Studies of Early Childhood Settings." In *Reasons to Take Stock and Strengthen Our Measures of Quality*, Martha Zaslow, ed. Baltimore, MD: Brookes Publishing.

Camilli, Gregory, Sadako Vargas, Sharon Ryan, and W. Steven Barnett. 2010. "Meta-Analysis of the Effects of Early Education Interventions on Cognitive and Social Development." *Teachers College Record* 112(3): 579–620.

Cascio, Elizabeth U., and Diane Whitmore Schanzenbach. 2013. "The Impacts of Expanding Access to High-Quality Preschool Education." NBER Working Paper No. 19735. Cambridge, MA: National Bureau of Economic Research.

Chetty, Raj, Jonathan N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane W. Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126(4): 1593–1660.

Chingos, Matthew M. 2015. *Breaking the Curve: Promises and Pitfalls of Using NAEP Data to Assess the State Role in Student Achievement*. Urban Institute Research Report. Washington, DC: Urban Institute.

Conley, Timothy, and Christopher Taber. 2011. "Inference with 'Difference-in-Differences' with a Small Number of Policy Changes." *Review of Economics and Statistics* 93(1): 113–125.

Correia, Sergio. 2014. "REGHDFE: Stata Module to Perform Linear or Instrumental-Variable Regression Absorbing any Number of High-Dimensional Fixed Effects." Statistical Software Components s457874. Boston: Boston College Department of Economics. Revised July 25, 2015.

Currie, Janet, and Duncan Thomas. 1995. "Does Head Start Make a Difference?" *American Economic Review* 85(3): 341–364.

Deming, David. 2009. "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start." *American Economic Journal: Applied Economics* 1(3): 111–134. doi:10.1257/app.1.3.111.

Donald, Stephen G., and Kevin Lang. 2007. "Inference with Difference-in-Differences and Other Panel Data." *Review of Economics and Statistics*. 89(2): 221–233.

Duckworth, Angela L., and David S. Yeager. 2015. "Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes." *Educational Researcher* 44(4): 237–251.

Duncan, Greg J., and Katherine Magnuson. 2013. "Investing in Preschool Programs." *Journal of Economic Perspectives* 27(2): 109–132.

Feller, Avi, Todd Grindal, Luke Miratrix, and Lindsay Page. 2014. "Compared to What? Variation in the Impact of Early Childhood Education by Alternative Care-Type Settings." Harvard University Working Paper. Cambridge, MA: Harvard University.

Fitzpatrick, Maria D. 2008. "Starting School at Four: The Effect of Universal Pre-Kindergarten on Children's Academic Achievement." *B.E. Journal of Economic Analysis & Policy* 8(1) Advances, Article 46: 1–38.

Garces, Eliana, Duncan Thomas, and Janet Currie. 2002. "Longer-Term Effects of Head Start." *American Economic Review* 92(4): 999–1012.

Gormley, William T., Jr., Deborah Phillips, and Ted Gayer. 2008. "Preschool Programs Can Boost School Readiness." *Science* 320(5884): 1723–1724. doi:10.1126/science.1156019.

Griliches, Zvi, and Jerry A. Hausman. 1986. "Errors in Variables in Panel Data." *Journal of Econometrics* 31: 93–118.

Grissmer, David, Ann Flanagan, Jennifer Kawata, and Stephanie Williamson. 2000. *Improving Student Achievement: What State NAEP Test Scores Tell Us*. Santa Monica, CA: RAND Corporation.

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rifkin. 2003. "Does Peer Ability Affect Student Achievement?" *Journal of Applied Econometrics* 18(5): 527–544.

Heckman, James J. 2015. "Quality Early Childhood Education: Enduring Benefits." October 15. http://heckmanequation.org/content/quality-early-childhood-education-enduring-benefits (accessed October 28, 2015).

Heckman, James J., Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103(6): 2052–2086.

Hoxby, Caroline. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." NBER Working Paper No. 7867. Cambridge, MA: National Bureau of Economic Research.

Kearney, Melissa S., and Phillip B. Levine. 2015. "Early Childhood Education by MOOC: Lessons from Sesame Street." NBER Working Paper No. 21229. Cambridge, MA: National Bureau of Economic Research.

Keys, Tran D., George Farkas, Margaret R. Burchinal, Greg J. Duncan, Deborah L. Vandell, Weilin Li, Erik A. Ruzek, and Carollee Howes. 2013. "Preschool Center Quality and School Readiness: Quality Effects and Variation by Demographic and Child Characteristics." *Child Development* 84(4): 1171–1190.

Kline, Patrick, and Christopher Walters. 2015. "Evaluating Public Programs with Close Substitutes: The Case of Head Start." NBER Working Paper No. 21658. Cambridge, MA: National Bureau of Economic Research.

Ladd, Helen F., Clara G. Muschkin, and Kenneth A. Dodge. 2014. "From Birth to School: Early Childhood Initiatives and Third-Grade Outcomes in North Carolina." *Journal of Policy Analysis and Management* 33(1): 162–187. doi:10.1002/pam.21734.

Lipsey, Mark W., Dale C. Farran, and Kerry G. Hofer. 2015a. "A Randomized Control Trial of a Statewide Voluntary Prekindergarten Program on Children's Skills and Behaviors through Third Grade." Vanderbilt University Working Paper. Nashville: Vanderbilt University.

Lipsey, Mark W., Christina Weiland, Hirokazu Yoshikawa, Sandra Jo Wilson, and Kerry G. Hofer. 2015b. "The Prekindergarten Age-Cutoff Regression-Discontinuity Design Methodological Issues and Implications for Application." *Educational Evaluation and Policy Analysis* 37(3): 296–313.

Ludwig, Jens, and Douglas L. Miller. 2007. "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design." *Quarterly Journal of Economics* 122(1): 159–208.

Magnuson, Katherine A., Christopher Ruhm, and Jane Waldfogel. 2007. "Does Prekindergarten Improve School Preparation and Performance?" *Economics of Education Review* 26(1): 33–51.

Minervino, Jim. 2014. *Lessons from Research and the Classroom: Implementing High-Quality Pre-K that Makes a Difference for Young Children*. Seattle, WA: Bill and Melinda Gates Foundation.

National Institute for Early Education Research (NIEER). 2003–2015. *State of Preschool Yearbooks*. New Brunswick, NJ: Rutgers University, National Institute for Early Education Research.

Neidell, Matthew, and Jane Waldfogel. 2010. "Cognitive and Noncognitive Peer Effects in Early Education." *Review of Economics and Statistics* 92(3): 562–576.

Phillips, Deborah A., William T. Gormley, and Amy E. Lowenstein. 2009. "Inside the Pre-Kindergarten Door: Classroom Climate and Instructional Time Allocation in Tulsa's Pre-K Programs." *Early Childhood Research Quarterly* 24(3): 213–228.

Reardon, Sean F., Demetra Kalogrides, and Andrew D. Ho. 2016. "Linking U.S. School District Test Score Distributions to a Common Scale, 2009–2013." Stanford CEPA Working Paper 16-09. Stanford: Stanford University, Center for Education Policy Analysis.

Rosinsky, Kristina L. 2014. "The Relationship Between Publicly Funded Preschool and Fourth Grade Math Test Scores: A State-Level Analysis." Master's thesis, Georgetown University.

Sabol, Terri J., S.L. Soliday Hong, R.C. Pianta, and M.R. Burchinal. 2013. "Can Rating Pre-K Programs Predict Children's Learning?" *Science* 341(23): 845–846.

Stevens, Katharine B., and Elizabeth English. 2016. *Does Pre-K Work? The Research on Ten Early Childhood Programs—And What It Tells Us*. Washington, DC: American Enterprise Institute.

Taylor, L.L., and W.J. Fowler, Jr. 2006. *A Comparable Wage Approach to Geographic Cost Adjustment*. Research and Development Report No. NCES-2006-321. Washington, DC: National Center for Education Statistics.

U.S. Department of Education. 2015a. *National Assessment of Educational Progress (NAEP), various years, 1990–2015 Mathematics Assessments*. Washington, DC: Institute of Education Sciences, National Center for Education Statistics.

———. 2015b. *The Condition of Education.* Washington, DC: U.S. Department of Education.

U.S. Department of Health and Human Services. 2010. *Head Start Impact Study* (Final Report). Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.

Weiland, Christina, and Hirokazu Yoshikawa. 2013. "Impacts of a Prekindergarten Program on Children's Mathematics, Language, Literacy, Executive Function, and Emotional Skills." *Child Development* 84(6): 2112–2130.

Zaslow, Martha, Rachel Anderson, Zakia Redd, Julia Wessel, Louisa Tarullo, and Margaret Burchinal. 2010. *Quality Dosage, Thresholds, and Features in Early Childhood Settings: A Review of the Literature, OPRE 2011-5*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

APPENDICES

These online appendices provide additional background literature review, additional information on the data, some checks for possible biases due to migration and endogeneity, and additional results for different samples and groupings.

APPENDIX A

LITERATURE REVIEW DETAILS

The review of the literature in the main text makes several statements about what the pre-K literature "shows." Table A1 describes the specific programs referenced.

Table A1 Summary of Literature on Effects of Pre-K over Various Horizons

Type of study	Study	Short-run (< 1 year)	Medium-run (primary school)	Long-run (high school+)
Classic experiments	Perry: 2 years of half-day pre-K, @$10,559 per student-year.	18 percentiles (ES=0.59)	3 percentiles at end of 3rd grade (ES=0.10), 1 percentile at end of 4th grade (ES=0.04). Reduces special ed. for mental impairment by 20 pp, overall special ed. by 5 pp. (ns). Reduces grade repetition by 5 pp, grade repetition by 2 or more years by 7 pp (ns).	19% earnings boost; 50–59% crime reduction; reduced smoking/drinking
	Abecedarian: Five years of full-time full-yr care/pre-K, birth to 5, @$17,856 per student-year.	19 percentiles (ES=0.50)	10 percentiles at 3rd grade (ES=0.27)	26% earnings boost; no crime effect; reduced risk factors for cardiovascular disease
Quasi-experiments	Chicago Child-Parent Center: 2 years of half-day pre-K, @$5,668 per student-year. Benefits did not increase much for 2-year vs. 1-year.	11 percentiles (ES=0.38)	3 percentiles at 3rd grade (ES=0.07); grade retention by age 15 drops by 15 pp; special ed. by age 18 drops by 10 pp.	8% earnings boost; 22% reduction in felony arrests; 26% reduction in depression, 24% reduction in substance abuse.
	Head Start-siblings (Deming): 1–2 yrs of mix of half-day versus full-day, although modal is 1-year, @$9,249 per student-year.	5 percentiles at ages 5-6 (ES=0.15)	4 percentiles at ages 7–10 (ES=0.13); 2 percentiles at ages 11–14 (ES=0.06). Reduced diagnosis of learning disability by 6 pp, ever grade repetition by 7 pp.	Predicted 11% earnings gain; no crime effect; percentage in poor health drop by 7 pp.
	Head Start-siblings (Currie & Thomas; Garces, Thomas, & Currie)	Currie & Thomas: 7 percentiles at age 5 (ES=0.21)	Currie and Thomas: 6 percentiles (ES=0.18) for whites, 0 for blacks. White reduction in any grade retention by age 10+ is 47 pp, 0 reduction for blacks.	Garces-Thomas-Currie: whites 28 pps more likely to complete high school, 28 pps more likely to attend college; 0 attainment effects for blacks. Blacks 13 pp less likely to be charged with crime, no white effects.
	Head Start (Ludwig & Miller) comparison across counties with different grant-writing assistance (geographic study).		Grant-writing assistance reduces Head Start preventable mortality at ages 5–9 by 30–50 percent. No effects on 8th grade test scores.	Grant-writing assistance increases high school completion and college attendance by 3 to 5 pp.

Type of study	Study	Short-run (< 1 year)	Medium-run (primary school)	Long-run (high school+)
Meta-analyses	(Duncan & Magnuson)	9 percentiles at end of program (ES=0.27)	5 percentiles by 4th grade (ES=0.15)	
	Camilli et al.;	14 percentiles at end of program (ES=0.39)	4 to 5 percentiles both at ages 5–10 and ages 10+ (ES=0.14–0.15)	
Other studies	Head Start Experiment	7 percentiles at end of program (ES=0.22)	2 percentiles at end of 3rd grade (ES=0.06)	
	RDD Barnett et al. studies of 7 states	11 percentiles at beginning of kindergarten (ES=0.31)		
	RDD Gormley, Phillips, and Gayer (Tulsa) and matching follow-up study. $5,304 for half-day pre-K for one school year, $10,608 for full-day pre-K.	RDD results: At kdg entrance, full-day has pctile gain of 18 for FRL students, 17 for non; half-day is 11 for FRL, 10 for non (ESs = 1.07, 0.96, 0.66, 0.58). PSM results appear to cut these ESs in half for reading, by 1/3 for math.	7 percentiles (ES=0.18) in math for late cohort, less than 0.4 pctiles (ES=0.01) for early cohort in math. In reading, 4 percentiles for late cohort (ES =0.09), minus 1 percentile for early cohort (ES=−0.03). Only late cohort math result is statistically significant.	
	RDD Weiland/Yoshikawa (Boston). Full-day pre-K program, cost of $15,000 to $17,000 per student.	21 percentiles gain at kindergarten entrance for FRL students (ES =0.59), 15 percentiles for non-FRL students (ES=0.38)		

Type of study	Study	Short-run (< 1 year)	Medium-run (primary school)	Long-run (high school+)
	Tennessee experiment (Lipsey et al.) Full-day 1-year program at $4,669 per student.	8 percentile gain at end of program (ES=0.24) based on comparison group. 17 percentile gain at kindergarten entry (ES=0.49) based on RDD.	3 percentile LOSS at end of 3rd grade (ES=−0.1).	
Kindergarten class quality	Chetty et al.: 1 standard deviation improvement in kindergarten class quality, as measured by end of kindergarten peer scores.	6 percentile gain at end of kindergarten (ES=0.16)	1 percentile gain at end of 4th grade (ES=0.03)	3% gain in adult earnings
Recent geographic studies	Fitzpatrick (Georgia): Georgia: $5,590 per student for full-day program.		6 percentile points (ES=0.15) for both math and reading NAEP scores at 4th grade; significant with clustered standard errors, insignificant with Conley-Taber corrections.	
	Cascio/Schanzenbach (Oklahoma/Georgia): OK: $7,782 per student for mix of half-day and full-day programs: GA: $5,590 per student for full-day program.		4th grade: FRL gain of 14 percentiles in both math & reading NAEP scores (ES=0.39, .40); non-FRL gain of 4 pctiles in math, loss of 6 pctiles in reading (ES=0.10, −0.16). 8th grade: FRL gain of 11 pctiles in math, 4 pctiles in reading (ES=0.33, 0.12); non-FRL loss of 5 pctiles in math, 4 pctiles in reading (ES=−0.12, -0.09). Only FRL 4th grade gains and 8th grade math gains are statistically significant in main reported estimates; none of estimates are statistically significant with Conley-Taber corrections.	
	Ladd, Muschkin, and Dodge (North Carolina) More at Four, a full-day pre-K program, @$6,143 per student.		20 percentiles in math (ES=0.54), 25 percentiles in reading for North Carolina tests (ES=0.66)	
	Rosinsky, panel data on all states		State funded pre-K reduces 4th grade math NAEP test scores by 6 percentiles for all students (ES=−0.14), and 7 percentiles for low-income students (ES=−0.26). All publicly funded pre-K reduces NAEP scores of all students by 5 percentiles (ES=−0.11), low-income students by 6 percentiles (ES=−0.20).	

APPENDIX B

MORE ON CCD AND NAEP DATA SOURCES AND HOW WE USE THEM IN ESTIMATION

Preparing the CCD and NAEP data—especially merging them—requires researcher judgment to overcome various issues with the data sets. This appendix section provides more detail on how we handle these data issues, and on what years and states are available for analysis.

PRE-K DATA IN CCD

The CCD provides enrollment for the universe of public schools in the United States. For state-level analyses, we take reported pre-K, both overall and by race, from the CCD's state-level files and divide by state-year estimates of the population of four-year-olds from the National Cancer Institute's SEER population data. For school-level analyses, we take reported pre-K and grade 1 enrollment from the CCD's school-level files and divide the former by the latter, top-coding the ratio at 1 if it exceeds 1 but is less than 1.5; we set to "missing" ratios that exceed 1.5. For district-level analyses, we again take reported pre-K and grade 1 enrollment from the CCD's school-level files, as grade-specific enrollment is not reported in the district-level files. We sum enrollments in each grade for all schools within a district, and then take the ratios of these sums, with the same top-coding rule applied. (At the district-level, only a few cases are top-coded, and a negligible number have ratios exceeding 1.5.) Although school-level pre-K enrollment by race of student is available in recent years, we do not use it given its limited availability.

Not every school or state reports a valid number for pre-K enrollment each year. In most of these cases, there is a missing code for "not applicable." That is, instead of entering a zero, the school or district reporting official indicated that the pre-K enrollment field was not applicable because there was no pre-K program. In some other cases, however, it appears that the state or school may have positive pre-K enrollment that is incorrectly reported as a true missing (different than the "not applicable" code). California, for example, never reports pre-K enrollment by school, or statewide by race, but does report positive pre-K enrollment for the state in the aggregate. We code the "not applicable" missings as zeros and the true missings as such, with the following two exceptions: 1) if a school or state reports positive pre-K enrollment in year $t-1$ and year $t+1$ but a "not applicable" in year t, we code it as a missing; 2) if positive pre-K was reported at the state level but no school in that state and year reported positive pre-K enrollment (i.e., California), all such schools were coded to missing that year.

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS (NAEP)

The NAEP, also known as the Nation's Report Card, is a nationally representative assessment periodically given to U.S. 4th graders, 8th graders, and 12th graders in several academic subjects. Mathematics and reading assessments have been given to representative samples of 4th graders and 8th graders in every state biennially since 2003; prior to that year, most states participated in the math and reading assessments, which were slightly less frequent. The NAEP is a multistage probability sample in which schools are selected for participation, and approximately 30 students in each school are given the survey instrument and assessment. While statistics at the state-by-demographic levels are released publicly, we employ the restricted-use version that contains individual-level data.

The restricted-use NAEP collects rich information about each student, school, and district (if applicable), some of which we use as described in the text. Particularly relevant for this paper, the NAEP since 1998 records an ID number for each participating school and district that allows these units to be matched longitudinally in successive waves (if they were resampled), as well as merged with additional CCD data, including pre-K enrollment and other characteristics, as noted above. We can match all but a trivial (< 0.2 percent) fraction of public school students in the NAEP to CCD schools and districts when identifiers are present.

YEARS AND STATES IN MERGED CCD PRE-K AND NAEP 4TH-GRADE TEST DATA

Due to the timing of the NAEP, and limitations in the pre-K data in the CCD, not all states and years are available for analysis. Tables B1 and B2 list all NAEP years and states for which we have math NAEP data (Table B1) or reading NAEP data (Table B2) at 4th grade, along with matching pre-K data five years earlier.

Note that in these tables, pre-K data are lagged five years from shown (NAEP) year. Also, the math NAEP was not conducted in 1998 and 2002; the reading NAEP was not conducted in 1996 and 2000.

Table B1 States and Years with Math NAEP Data and Valid Pre-K Measures

State FIPS code	1996	1998	2000	2002	2003	2005	2007	2009	2011	2013
Alabama								X	X	X
Alaska					X	X	X	X	X	X
Arizona			X		X	X	X	X	X	X
Arkansas			X		X	X	X	X	X	X
California										
Colorado					X	X	X	X	X	X
Connecticut			X		X	X	X	X	X	X
Delaware					X	X	X	X	X	X
DC			X		X	X	X	X	X	X
Florida					X	X	X	X	X	X
Georgia			X		X	X	X	X	X	X
Hawaii			X		X	X	X	X	X	X
Idaho								X	X	X
Illinois			X		X	X	X	X	X	X
Indiana			X		X	X	X	X	X	X
Iowa			X		X	X	X	X	X	X
Kansas			X			X	X	X	X	X
Kentucky						X	X			X
Louisiana			X		X	X	X	X	X	X
Maine			X		X	X	X	X	X	X
Maryland			X		X	X	X	X	X	X
Massachusetts			X		X	X	X	X	X	X
Michigan			X		X	X	X	X	X	X
Minnesota			X		X	X	X	X	X	X
Mississippi			X		X	X	X	X	X	X
Missouri			X		X	X	X	X	X	X
Montana			X		X	X	X	X	X	X
Nebraska			X		X	X	X	X	X	X
Nevada			X		X	X	X	X	X	X
New Hampshire					X	X	X	X	X	X
New Jersey							X	X	X	X
New Mexico			X		X	X	X	X	X	X
New York			X		X	X	X	X	X	X
North Carolina			X			X	X	X	X	X
North Dakota					X	X	X	X	X	X
Ohio			X		X	X	X	X	X	X
Oklahoma			X		X	X	X	X	X	X
Oregon			X		X	X	X	X	X	X
Pennsylvania					X	X	X	X	X	X
Rhode Island			X		X	X	X	X	X	X
South Carolina					X	X	X	X	X	X
South Dakota					X	X	X	X	X	X
Tennessee									X	X
Texas			X		X	X	X	X	X	X
Utah			X		X	X	X	X	X	X
Vermont			X		X	X	X	X	X	X
Virginia			X		X	X	X	X	X	X
Washington					X	X	X	X	X	X
West Virginia			X		X	X	X	X	X	X
Wisconsin			X		X	X	X	X	X	X
Wyoming						X		X	X	X

Table B2 States and Years with Reading NAEP Data and Valid Pre-K Measures

State FIPS code	1996	1998	2000	2002	2003	2005	2007	2009	2011	2013
Alabama								X	X	X
Alaska				X	X	X	X	X	X	X
Arizona		X		X	X	X	X	X	X	X
Arkansas		X		X	X	X	X	X	X	X
California										
Colorado		X			X	X	X	X	X	X
Connecticut		X		X	X	X	X	X	X	X
Delaware		X		X	X	X	X	X	X	X
DC		X		X	X	X	X	X	X	X
Florida		X		X	X	X	X	X	X	X
Georgia		X		X	X	X	X	X	X	X
Hawaii		X		X	X	X	X	X	X	X
Idaho								X	X	X
Illinois		X		X	X	X	X	X	X	X
Indiana				X	X	X	X	X	X	X
Iowa		X		X	X	X	X	X	X	X
Kansas		X		X		X	X	X	X	X
Kentucky						X	X			X
Louisiana		X		X	X	X	X	X	X	X
Maine				X	X	X	X	X	X	X
Maryland		X		X	X	X	X	X	X	X
Massachusetts		X		X	X	X	X	X	X	X
Michigan		X		X	X	X	X	X	X	X
Minnesota		X		X	X	X	X	X	X	X
Mississippi		X		X	X	X	X	X	X	X
Missouri				X	X	X	X	X	X	X
Montana		X		X	X	X	X	X	X	X
Nebraska				X	X	X	X	X	X	X
Nevada		X		X	X	X	X	X	X	X
New Hampshire		X		X	X	X	X	X	X	X
New Jersey							X	X	X	X
New Mexico		X		X	X	X	X	X	X	X
New York		X		X	X	X	X	X	X	X
North Carolina		X		X		X	X	X	X	X
North Dakota				X	X	X	X	X	X	X
Ohio				X	X	X	X	X	X	X
Oklahoma		X		X	X	X	X	X	X	X
Oregon		X		X	X	X	X	X	X	X
Pennsylvania				X	X	X	X	X	X	X
Rhode Island		X		X	X	X	X	X	X	X
South Carolina						X	X	X	X	X
South Dakota				X	X	X	X	X	X	X
Tennessee									X	X
Texas		X		X	X	X	X	X	X	X
Utah		X		X	X	X	X	X	X	X
Vermont				X	X	X	X	X	X	X
Virginia		X		X	X	X	X	X	X	X
Washington		X		X	X	X	X	X	X	X
West Virginia		X		X	X	X	X	X	X	X
Wisconsin		X		X	X	X	X	X	X	X
Wyoming						X		X	X	X

APPENDIX C

TESTS FOR ENDOGENEITY BIAS DUE TO EFFECTS OF LAGGED PRE-K TEST SCORES ON PRE-K ENROLLMENT RATES

As mentioned in the text, one concern in our "natural experiment" is that pre-K enrollment rates are not (as good as) randomly assigned, but rather chosen by school districts based on many factors. If past test scores are among the variables that "cause" current pre-K enrollment rates, and are also extremely persistent, then any observed correlation between current pre-K enrollment and future test scores may in some part reflect the effect of past test scores on current pre-K enrollment rates.

To test for this problem, we performed Granger-causality tests using our merged CCD/NAEP pre-K and test score database. Specifically, we regressed pre-K enrollment rates on lagged test scores as well as their own lagged values. If the first relationship is statistically significant, then past test scores are in some sense potentially "causal" to current pre-K enrollment rates. These regressions also controlled for other potential explanatory variables, including year fixed effects, district fixed effects, and time-varying district-specific controls (public school spending, Head Start, and private pre-K options). We also tested whether the effect of lagged test scores on current pre-K enrollment rates varied between quality states and all other states.

We included specifications that used a one-period lag in the test score and pre-K variables, as well as both one- and two-period lags. Because of the nature of our data, and specifically the fact that the data are not available every year, the first lag is usually two years in length, and sometimes three years; the second lag is typically four years ago, and sometimes five years ago.

In general, the results show that lagged test scores are not statistically significant in explaining current pre-K enrollment rates. Therefore, test scores do not significantly “Granger-cause” district pre-K enrollment rates. The sizes of the estimated effects are also substantively modest.

Table C1 illustrates our results in the case of math test scores. When we include only one lag, past math NAEP scores are not significantly correlated with current district pre-K enrollment rates. A test of the proposition that all the lagged test score coefficients are zero yields a p-value of 0.321. When we add a second lag, the lagged effects have marginally stronger correlation, but the p-value of 0.193 still implies statistical insignificance. The “standardized” effects are also relatively small: a 1 standard deviation in test scores is associated with no more than a 0.25 standard deviation change in pre-K enrollment rates.

Table C1 "Effects" of Lagged Math Test Scores on District Pre-K Enrollment Rates

	(1)	(2)
Lagged math test score	−0.000144	−0.000208
[standard error in brackets]	[0.000357]	[0.000613]
2nd lag in lagged math test score		−0.000726
		[0.000467]
Lagged math test score interacted with quality state indicator	0.00226	0.00455
	[0.00215]	[0.00317]
2nd lag in math test score interacted with quality state		0.00274
		[0.00194]
Lagged district pre-K proportion	0.249***	0.169**
	[0.0395]	[0.0675]
2nd lag in district pre-K proportion		−0.103***
		[0.0293]
Number of observations	7,500	4,400
Adjusted R-squared	0.886	0.908
Mean (adj) math test score in sample	−1.669	−1.590
Math test score standard deviation	9.274	9.070
Pre-K proportion mean	0.244	0.254
Pre-K proportion standard deviation	0.263	0.262
p-value for all test score variables in regression	0.321	0.193
Standardized effect of test score on pre-K in nonquality states (st. dev. units)	−0.005	−0.032
Standardized effect of test score in quality states	0.075	0.220

NOTE: This table reports two different regressions, with district-years as observations (rounded to the nearest 10 to comply with disclosure restrictions), and with the dependent variable in both cases being the district pre-K enrollment rate. The first column of numbers shows coefficient estimates (and standard errors, in brackets) with a single lag of test scores and pre-K enrollment rates on the right-hand side. The second column contains two lags in test scores and enrollment rates. As can be seen, including two lags significantly reduces the number of observations. Both regressions include year fixed effects, district fixed effects, and time-varying district controls. Some descriptive statistics are included for the test score and pre-K enrollment rate variables. Test scores are measured in percentiles, but are adjusted for student characteristics as described in the main text, and recentered from the national mean for each year. The last two rows report the effects of a one-standard deviation change in test scores on the pre-K enrollment rate measured in standard deviation units.

APPENDIX D

TESTS FOR ATTENUATION BIAS DUE TO MIGRATION

As mentioned in the text, one concern about our estimates is that there may be migration into and out of the geographic unit between the year that children would be in pre-K and the year they are tested for NAEP in 4th grade (or 8th grade, for some of the appendix results reported). This bias is a problem only if we think there are no positive spillover effects of an individual child's skills on other children, and if our goal is to estimate the effects of pre-K on each *individual* child's test scores and other outcomes. If we are instead simply interested in the overall effects of a geographic unit's pre-K enrollment rates on its subsequent test scores—effects which may be reduced by in-migration and out-migration, but increased by skill spillovers—then the results including migration effects may be perfectly satisfactory.

However, if we are interested in individual child effects, and skill spillovers are small and migration is random, then in-migration and out-migration will tend to attenuate the effects of pre-K. Under such attenuation, the true effect of pre-K on the individual child is equal to our estimated effect divided by (1 − migration rate).

To address this possibility, we examined approximated school district migration rates between the "pre-K year" and the "grade 4 year" using confidential geographic data from the Panel Study of Income Dynamics (PSID). These PSID data identify the location of individual persons down to the census tract level. We operationalized the "pre-K year" as ages 4 or 5 and the "grade 4 year" as ages 9 or 10, and we examined similar birth cohorts as used in the NAEP/CCD analysis (1988–2000).

Because census tracts do not perfectly match to school districts, we used multiple procedures for the matching assignment. In cases where census tracts were wholly contained within a school district, the assignment was straightforward. For cases where tracts split across school districts, we assigned a tract to a school district if at least x% of the tract's population in 2000 resided in the school district. Otherwise, the census tract was not assigned to that school district.[41] We varied the cutoff threshold x between 50% and 100%. As we increase the cutoff, we tend to include relatively more individuals in larger school districts in our sample, and relatively fewer individuals in smaller school districts, as larger school districts tend to be of sufficient size that a larger proportion of census tracts fall within the district boundaries.

Once we have defined a "pseudo-school district" using this procedure, we then measure the likelihood that an individual moved school districts over the five-year period between ages 4–5 and ages 9–10. We calculated these measures for all children, and for groups defined by race and by free and reduced-price lunch status. We also examined how "pseudo-district" migration rates varied between early (1988–1993) and later (1994–2000) birth cohorts.

Table D1 reports the results. The main take-away for our purposes is that the overall out-migration rate is modest. If one looks at the bottom row, the average out-migration rate varies between 17 percent and 22 percent, depending on how tightly one defines these "pseudo-districts." If migration is random and there are no spillovers, this implies that individual student effects are 20–28 percent higher than our reported estimates.

[41] We used the University of Missouri Data Center (http://mcdc2.missouri.edu/websas/geocorr2k.html) for the mapping between census tracts and school districts.

Table D1 Migration Rates from "School Districts" between Pre-K and Grade 4

		Migration rates (%), different sensitivity thresholds for tract-district matching, x					
Group	Obs.	50	60	70	80	90	100
Non-FRL, early birth cohort (1988–1993)	835	20.0	19.0	18.5	18.2	17.5	16.7
Non-FRL, late birth cohort (1994–2000)	964	21.3	19.8	18.3	17.5	16.0	14.9
FRL, first half birth cohort (1988–1993)	636	23.1	23.0	22.7	21.4	21.2	19.6
FRL, second half birth cohort (1994–2000)	631	29.1	27.6	26.3	26.0	26.1	21.2
Non-FRL	1,799	20.7	19.4	18.4	17.8	16.7	15.7
FRL	1,267	25.9	25.2	24.3	23.6	23.5	20.3
White, early birth cohort (1988–1993)	757	22.8	22.1	21.7	21.1	20.4	19.7
White, late birth cohort (1994–2000)	859	23.4	21.8	20.6	19.8	18.7	17.3
Minority, early birth cohort (1988–1993)	684	17.3	16.7	16.1	15.8	15.4	14.1
Minority, late birth cohort (1994–2000)	724	24.9	23.8	21.0	21.0	20.4	15.4
White	1,616	23.1	21.9	21.1	20.4	19.5	18.4
Minority	1,408	21.1	20.2	18.5	18.4	17.8	14.7
Early birth cohort (1988–1993)	1,478	21.1	20.5	20.0	19.4	18.8	17.8
Late birth cohort (1994–2000)	1,610	23.3	21.8	20.4	19.8	18.7	16.5
All	3,088	22.3	21.2	20.2	19.6	18.8	17.1

NOTE: Authors' tabulations of confidential geographic PSID microdata. The number of observations is the maximum number of potential observations for indicated group.

We also find some interesting patterns in these out-migration rates. Notably, they tend to be higher for the later birth cohort. This contradicts conventional wisdom about the decline of migration in the United States, although the pattern holds for a particular geographic level that could range from neighborhood to county. The pattern could differ at lower (moving blocks) or higher (metropolitan area or state) levels. It is also of interest that the increase in "district" out-migration rates is higher for minority and lower-income groups than for more advantaged groups. Perhaps the increase in out-migration reflects gentrification that is more prominent in the later birth cohort. We leave these hypotheses for future research.

The key point is that even for the different subgroups and birth cohorts, migration rates never are as great as 30 percent. Therefore, attenuation bias is likely limited.

What if migration is not random? That is, what if net migration is selected in a way that could significantly increase or decrease test scores due to compositional changes? For this possibility to bias our estimates, any net migration effects on test scores would have to systematically increase or decrease as the pre-K proportion in the district goes from 0 percent enrollment to 100 percent enrollment. Although in-migrants and out-migrants may differ, it seems unlikely that the relative composition of the net migration flow would change so dramatically with pre-K enrollment rates.

APPENDIX E

ROBUSTNESS CHECKS: HOW RESULTS VARY FOR DIFFERENT SPECIFICATIONS

This appendix reports various types of robustness checks. We consider how our results change under different econometric specifications, with different groups examined, when we adopt school-level analyses, and for 8th grade (rather than 4th grade) outcomes.

SENSITIVITY TO DIFFERENT APPROACHES FOR DEALING WITH TIME TRENDS

We first consider different ways of dealing with time-period effects. Table E1 shows how the state-level results (Table 4) change when we add state-specific linear time trends. As can be seen, adding time trends does not appreciably change the results, although it does tend to reduce precision. Standard errors increase between 10 and 30 percent.

Table E1: The Effects of Pre-K on State-Level 4th Grade Outcomes: FRL vs Non-FRL students (including state-specific linear time trends)

	(1)	(2)
	FRL students	Non-FRL students
Math scores (percentile)	2.069	3.421
	(2.443)	(3.725)
Reading scores (percentile)	1.259	0.233
	(1.724)	(2.703)
Special education (proportion x100)	0.00	0.70
	(2.27)	(1.34)
Over-age for grade (proportion x100)	2.60	2.56
	(4.27)	(2.22)

NOTE: See Table 4.

Table E2 takes the overall district-level outcomes (Table 5) and adds a complete set of dummies for each state-year cell. The identification in this specification comes from within-district variation in pre-K relative to the constituent state's own flexible trend in pre-K, which is rather demanding. Results are nonetheless robust to this specification, although there now appear to be modest, but statistically significant, reductions in special education assignment due to pre-K.

Table E2: The Effects of Pre-K on District-Level 4th Grade Outcomes (including state-by-year dummies)

	(1)	(2)
Math scores (percentile)	−0.258	−0.172
	(0.539)	(0.530)
Reading scores (percentile)	−0.519	−0.358
	(0.526)	(0.501)
Special education (proportion ×100)	−1.10**	−1.16**
	(0.43)	(0.43)
Over-age for grade (proportion ×100)	−0.17	−0.14
	(0.40)	(0.39)
Include district fixed effects	Yes	Yes
Include district time-varying controls?	No	Yes
Include state-by-year fixed effects?	Yes	Yes

NOTE: See Table 5.

One concern with controlling for unobserved heterogeneity with district fixed effects is that year-to-year variation in pre-K enrollment could be dominated by measurement error or other noise, which will attenuate the coefficient estimates (Griliches and Hausman 1986). One approach to address this issue is to "long-difference" the data and look at within-district changes over the sample horizon. Table E3 uses this approach, examining the change in district outcomes between the last and first observed years in the data (controlling for the number of years between the two). These results are quite similar to the analogous results in Table 5, but with slightly greater imprecision.

Table E3: The Effects of Pre-K on District-Level 4th Grade Outcomes (long differences instead of fixed effects)

	(1)	(2)
Math scores (percentile)	−0.168	−0.206
	(0.730)	(0.697)
Reading scores (percentile)	−2.195***	−1.858***
	(0.718)	(0.679)
Special education (proportion ×100)	−0.24	−0.44
	(0.60)	(0.59)
Over-age for grade (proportion ×100)	−0.11	−0.20
	(0.55)	(0.54)
Include long differences	Yes	Yes
Include (changes in) district time-varying controls?	No	Yes

NOTE: * significant at the 0.10 level; ** significant at the 0.05 level; *** significant at the 0.01 level. Each cell is from a separate regression of the *change* in outcome on the *change* in pre-K measure, across districts that appear at least twice in the NAEP data. The change represents the difference between the earliest and latest observation across NAEP years. The regressions also control for the (categorical) number of years elapsed between the change for each district, as this varies, as well as changes in the district's time-varying controls as described in the note to Table 4. Each observation is a district, and there are 5,250 observations for math scores; 5,490 for reading scores; 5,760 for special education; and 5,740 for over-age for grade. All observation counts have been rounded to the nearest 10 to comply with disclosure restrictions. Standard errors in parentheses are robust to heteroskedasticity.

Table E4 also examines the impact of specifying the model in long differences rather than as a panel model with fixed effects, but this time for the specification (Table 6) that allows pre-K effects to differ between "quality states" and other states. As before, the results are broadly similar to baseline.

Table E4: The Effects of Pre-K on District-Level 4th Grade Outcomes, Quality States vs. Other States (long differences instead of fixed effects)

	Other states	Quality states	Difference
Math scores (percentile)	−0.864	2.046	2.910*
	(0.776)	(1.337)	(1.499)
Reading scores (percentile)	−2.189***	−0.818	1.371
	(0.756)	(1.328)	(1.474)
Special education (proportion ×100)	−0.74	0.47	1.20
	(0.65)	(1.12)	(1.25)
Over-age for grade (proportion ×100)	−0.01	−0.72	−0.70
	(0.58)	(1.18)	(1.28)
Include long differences		Yes	
Include (changes in) district time-varying controls?		Yes	

NOTE: * significant at the 0.10 level; ** significant at the 0.05 level; *** significant at the 0.01 level. Each cell is from a separate regression of the *change* in outcome on the *change* in pre-K measure, across districts that appear at least twice in the NAEP data. The change represents the difference between the earliest and latest observation across NAEP years. The regressions also control for the (categorical) number of years elapsed between the change for each district, as this varies, as well as changes in the district's time-varying controls as described in the note to Table 5. Each observation is a district, and there are 5,250 observations for math scores; 5,490 for reading scores; 5,760 for special education; and 5,740 for over-age for grade. All observation counts have been rounded to the nearest 10 to comply with disclosure restrictions. Standard errors in parentheses are robust to heteroskedasticity.

RESULTS BY DIFFERENT GROUPINGS OF DISTRICTS

The text reported estimates when districts were grouped by racial composition. We now consider two alternative groupings.

Table E5 considers results when instead districts are grouped by the percentage of students eligible for a free or reduced-price lunch.

The results show that for both low-income and non-low-income districts, pre-K in "quality states" seems to raise math test scores. In contrast, pre-K in non-quality states has no significant positive effects on test scores in either type of district. In non-low-income districts, pre-K does tend to significantly reduce special education assignments in non-quality states, while increasing special education assignments in quality states.[42]

[42] As discussed in the main text, this could occur due to a greater likelihood of diagnosis.

Table E5: The Effects of Pre-K on District-Level 4th Grade Outcomes, Quality States vs. Other States: by District Income Composition

	Pooled	Quality Interaction		
		Other states	Quality states	Difference
Panel A: Districts with FRL Share <40%				
Math scores (percentile)	0.910	0.247	4.612**	4.365**
	(0.802)	(0.876)	(1.806)	(1.946)
Reading scores (percentile)	−0.431	−0.865	1.727	2.592
	(0.848)	(0.897)	(2.456)	(2.580)
Special education (proportion ×100)	−0.68	−1.39**	2.94*	4.33***
	(0.58)	(0.58)	(1.57)	(1.66)
Over-age for grade (proportion ×100)	0.24	0.53	−1.04	−1.57
	(0.44)	(0.42)	(1.30)	(1.34)
Panel B: Districts with FRL Share ≥40%				
Math scores (percentile)	0.157	−0.497	2.547*	3.044
	(0.841)	(1.060)	(1.529)	(1.976)
Reading scores (percentile)	−1.541**	−2.156***	0.471	2.627
	(0.690)	(0.823)	(1.618)	(1.927)
Special education (proportion ×100)	−1.25	−1.05	−1.89	−0.84
	(0.88)	(1.07)	(1.22)	(1.58)
Over-age for grade (proportion ×100)	−0.97	−0.68	−1.83	−1.15
	(0.65)	(0.70)	(1.39)	(1.51)

NOTE: * significant at the 0.10 level; ** significant at the 0.05 level; *** significant at the 0.01 level. See notes to Tables 5 and 6. Each panel represents regressions on the indicated subset of the data. District free and reduced-price lunch (FRL) shares are determined by the sample horizon average in the CCD. All regressions include district fixed effects and district time-varying controls. The number of district-year observations—all rounded to the nearest 10 to comply with disclosure restrictions—for panel A is: math, 10,370; reading, 11,500; special ed., 12,530; over-age, 12,580. The corresponding sample sizes for panel B are: 8,930; 9,910; 10,860; and 10,700. The number of unique districts is approximately between one-fourth and one-third the number of observations.

Therefore, these results by district income status contrast somewhat with the results by district racial composition. For the latter, we find evidence that in majority African-American districts pre-K increased test scores in both quality and non-quality states. In contrast, there is no evidence in the district income results that more "disadvantaged" districts show positive effects in all states.

Table E6 considers results when districts are grouped by district size.

Table E6: The Effects of Pre-K on District-Level 4th Grade Outcomes, Quality States vs. Other States: by District Size

	Pooled	Quality Interaction		
		Other states	Quality states	Difference
Panel A: Districts with <2,500 students				
Math scores (percentile)	−1.159*	−1.266*	−0.491	0.775
	(0.676)	(0.735)	(1.480)	(1.613)
Reading scores (percentile)	−1.043*	−1.306**	0.164	1.470
	(0.600)	(0.662)	(1.332)	(1.457)
Special education (proportion ×100)	−0.75	−0.89	−0.12	0.77
	(0.51)	(0.55)	(1.30)	(1.40)
Over-age for grade (proportion ×100)	1.22***	1.23**	1.14	−0.09
	(0.47)	(0.50)	(1.19)	(1.27)
Panel B: Districts with ≥2,500 students				
Math scores (percentile)	1.060	0.192	4.065***	3.873**
	(0.949)	(1.157)	(1.558)	(1.945)
Reading scores (percentile)	−1.319	−2.063**	1.229	3.292
	(0.845)	(0.960)	(1.764)	(2.006)
Special education (proportion ×100)	−0.96	−1.07	−0.56	0.51
	(0.95)	(1.13)	(1.33)	(1.69)
Over-age for grade (proportion ×100)	−1.41**	−0.98	−2.73**	−1.75
	(0.60)	(0.66)	(1.35)	(1.47)

NOTE: * significant at the 0.10 level; ** significant at the 0.05 level; *** significant at the 0.01 level. See notes to Tables 5 and 6. Each panel represents regressions on the indicated subset of the data. District enrollment sizes are determined by the sample horizon average in the CCD. All regressions include district fixed effects and district time-varying controls. The number of district-year observations—all rounded to the nearest 10 to comply with disclosure restrictions—for panel A is: math, 9,130; reading, 9,920; special ed., 10,830; over-age, 10,850. The corresponding sample sizes for panel B are: 10,190; 11,530; 12,620; and 12,470. The number of unique districts is approximately between one-fourth and one-third the number of observations.

The Table E6 results suggest that the text's finding that pre-K increases math test scores in "quality states" is largely driven by larger school districts. For these districts pre-K also reduces the proportion of students who are retained or otherwise "over-age" for grade. This pattern might be explained in several ways: the pre-K data may be measured more accurately for larger districts (due to less migration or data issues); pre-K may be more compensatory in larger school districts; or pre-K may be run more effectively in larger school districts.

RESULTS AT THE SCHOOL LEVEL

Our data also permit estimation of pre-K effects at the school level. We place less emphasis on school-level results because they are more susceptible to biases due to mismatch between who participates in pre-K at a particular school and who is tested in 4th grade. First, in-migration and out-migration at the school level is almost certainly higher than at the district level. Second, some school districts concentrate pre-K enrollment, or particular types of pre-K enrollment (e.g., full-day programs), in a subset of the district's schools. Despite these shortcomings, we present school-level results for completeness.

Appendix Table E7 presents descriptive statistics for the school-level samples.

Table E7: Summary Statistics for School-Level Samples, Grade 4

Variable	All schools		Majority-black schools		Majority-Hispanic schools		90%+ white schools	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Pre-K share	0.160	0.264	0.305	0.323	0.295	0.338	0.120	0.235
Share in "quality" states	0.111	0.314	0.195	0.396	0.084	0.277	0.043	0.203
Math percentile score, raw	45.7	15.6	27.0	11.7	34.9	11.4	53.3	12.1
Math percentile score, adjusted	−1.1	10.5	−6.0	9.1	−2.0	9.7	6.2	10.1
Reading percentile score, raw	46.8	14.6	30.0	10.9	34.8	11.1	54.5	10.8
Reading percentile score, adjusted	−0.9	10.0	−6.2	8.7	−3.0	8.9	0.9	9.5
Special ed. share (×100), raw	14.1	8.3	14.3	9.2	13.3	8.0	14.2	8.6
Special ed. share, adjusted	0.8	7.9	−2.5	8.8	−2.1	7.8	2.2	7.8
Over-age for grade, raw	9.9	8.8	17.3	10.7	12.9	8.9	8.3	7.7
Over-age for grade, adjusted	0.2	7.9	4.4	9.9	0.2	8.5	0.3	7.3
District per-pupil spending (000s)	7.4	1.9	7.8	1.7	7.0	1.8	7.7	2.0
N (school-years)	33,250		3,940		2,900		10,410	
Unique schools	10,810		1,280		960		3,190	

NOTE: All statistics are weighted by the number of NAEP students contributing to the relevant cell; unweighted statistics are similar. "Raw" statistics shown are as calculated in the NAEP data; adjusted statistics (used in the analyses) are recentered to have a weighted mean of 0 in each test year; see text for details. "Quality" states include MD, MA, NJ, NC, and OK. Per-pupil school spending is taken from the CCD and adjusted for inflation (to $1999) and the comparable wage index (across districts) via Taylor and Fowler (2006). The number of observations and unique schools vary slightly across outcomes; the statistics shown are the maximum across outcomes. Average cell sizes in the first stage—the number of students contributing to the outcome mean at the school-year level for each group—range between 10 and 30 across schools and years; the average is close to 20. All sample sizes are rounded to the nearest 10 to accord with disclosure restrictions.

As one might expect, the means in Table E7 are similar to the analogous district-level table in text Table 3. However, the variation in pre-K share and the various outcome measures tends to be higher at the school level. While this could lead to greater precision in estimation, it may give rise to more bias due to the mismatch between pre-K enrollment and subsequent outcomes.

Table E8 shows overall effects of pre-K on school-level 4th grade outcomes.

Table E8: The Effects of Pre-K on School-Level 4th Grade Outcomes

	(1)	(2)
Math scores (percentile)	0.093	0.073
	(0.452)	(0.445)
Reading scores (percentile)	−0.526	−0.444
	(0.393)	(0.390)
Special education (proportion ×100)	0.48	0.49
	(0.37)	(0.36)
Over-age for grade (proportion ×100)	−0.04	−0.05
	(0.34)	(0.34)
Include school fixed effects	Yes	Yes
Include school time-varying controls?	No	Yes

NOTE: * significant at the 0.10 level; ** significant at the 0.05 level; *** significant at the 0.01 level. Each cell is from a separate regression of the outcome on the pre-K measure, a set of test year dummies, a quartic in cost-of-living-adjusted, district-level current spending per student (averaged over the test year and preceding four years to account for time since pre-K), and the other controls as shown. School time-varying controls include: categorical dummies for the share of students eligible for free or reduced-price lunch, the student enrollment (size) of the school, the share of instructional staff working part-time, the share of students who are black in the school, the share of students who are Hispanic in the school, the number of private school pre-K slots available within 5 km of the school (normalized by the school's grade 1 enrollment), and the number of Head Start four-year old slots available within 10 km of the school (normalized by the school's grade 1 enrollment). Each observation is a school-year, and there are 26,930 observations (9,130 unique schools) for math scores; 30,150 (9,990) for reading scores; 33,250 (10,810) for special education; and 32,640 (10,620) for over-age for grade. All observation and school counts have been rounded to the nearest 10 to comply with disclosure restrictions. Standard errors in parentheses are clustered by district. The underlying dependent variables are school-year cell means that have been regression-adjusted for individual student characteristics and recentered so that the national weighted mean is zero for each year; see text for details. The independent variable is the ratio of pre-K enrollment in that school-year to first grade enrollment in that same school-year, taken from the CCD. The coefficients thus reflect the estimated effect of moving from 0 to 100 percent enrollment in pre-K.

As is true in the analogous district table (Table 5), there is no strong evidence of any positive effects of pre-K on student outcomes overall for all states, and in fact we can rule out large effects from a benefit-cost standpoint.

Table E9 reports school-level results broken down by quality states versus non-quality states.

Table E9: The Effects of Pre-K on School-Level 4th Grade Outcomes, Quality States vs. Other States

	Other states	Quality states	Difference
Math scores (percentile)	−0.094	0.990	1.084
	(0.502)	(1.159)	(1.303)
Reading scores (percentile)	−0.736	1.118	1.854
	(0.656)	(1.219)	(1.326)
Special education (proportion ×100)	0.49	0.53	0.04
	(0.39)	(0.90)	(0.96)
Over-age for grade (proportion ×100)	0.26	−1.70*	−1.96**
	(0.33)	(0.93)	(0.97)
Include school fixed effects		Yes	
Include school time-varying controls?		Yes	

NOTE: * significant at the 0.10 level; ** significant at the 0.05 level; *** significant at the 0.01 level. Each row is from a separate regression of the outcome on the pre-K measure, a set of test year dummies, a quartic in cost-of-living-adjusted, district-level current spending per student (averaged over the test year and preceding four years to account for time since pre-K), and school time-varying controls (see note to Table E8). The coefficients across columns show the pre-K measure, its interaction with an indicator variable for being a "quality program state" (equal to 1 for MD, MA, NJ, NC, and OK), and the net effect of pre-K in quality states. The estimates in column (1) thus shows the impact of moving from 0 to 100 percent enrollment in pre-K in all but the "quality states;" the estimates in column (2) shows the impact of moving from 0 to 100 percent enrollment in pre-K in the "quality states;" and the estimates in column (3) show the difference in these impacts. Each observation is a school-year; for sample sizes, see note to Table E8.

At the school level, the only statistically significant pre-K effect is the (marginal) reduction in the proportion over-age for grade in quality states. While the positive effect on test scores for quality states found in the district-level analysis remains, it is no longer statistically significant in school-level analyses, quite possibly a result of the bias issues discussed above.

Table E10 reports school-level results broken down by school racial composition.

Table E10: The Effects of Pre-K on School-Level 4th Grade Outcomes, Quality States vs. Other States: by School Racial Composition

	Pooled	Quality Interaction		
		Other states	Quality states	Difference
Panel A: Majority Black Schools				
Math scores (percentile)	0.421	0.267	1.197	0.930
	(1.127)	(1.161)	(2.574)	(2.577)
Reading scores (percentile)	−0.069	−0.740	3.685**	4.425**
	(0.826)	(0.942)	(1.846)	(1.951)
Special education (proportion ×100)	0.78	1.38	−2.39**	−3.77***
	(0.80)	(0.86)	(1.04)	(1.30)
Over-age for grade (proportion ×100)	−0.80	−0.62	−1.70	−1.08
	(1.03)	(1.11)	(2.06)	(2.22)
Panel B: Majority Hispanic Schools				
Math scores (percentile)	1.672	1.240	3.244	2.004
	(1.164)	(1.230)	(3.228)	(3.569)
Reading scores (percentile)	−0.053	−1.435	5.458*	6.893*
	(1.381)	(1.537)	(3.076)	(3.602)
Special education (proportion ×100)	0.49	0.24	1.49	1.25
	(1.16)	(1.34)	(1.84)	(2.23)
Over-age for grade (proportion ×100)	0.57	0.48	1.01	0.53
	(0.87)	(0.71)	(3.66)	(3.71)
Panel C: 90%+ White Schools				
Math scores (percentile)	−0.426	−0.571	4.497**	5.068
	(0.792)	(0.810)	(2.291)	(2.396)
Reading scores (percentile)	−0.436	−0.464	1.161	1.625
	(0.625)	(0.630)	(4.881)	(4.919)
Special education (proportion ×100)	−0.40	−0.29	−4.94	−4.65
	(0.64)	(0.65)	(4.04)	(4.07)
Over-age for grade (proportion ×100)	0.44	0.42	1.40	0.98
	(0.42)	(0.43)	(2.79)	(2.82)

NOTE: * significant at the 0.10 level; ** significant at the 0.05 level; *** significant at the 0.01 level. See notes to Tables E8 and E9. Each panel represents regressions on the indicated subset of the data. All regressions include school fixed effects and school time-varying controls. The number of school-year observations—all rounded to the nearest 10 to comply with disclosure restrictions—for panel A is: math, 3,100; reading, 3,580; special ed, 3,940; over-age, 3,900. The corresponding sample sizes for majority Hispanic districts are: 2,380, 2,700, 2,900, and 2,850. For 90% plus white districts: 8,560, 9,350, 10,410, and 10,340. The number of unique schools is approximately between one-fourth and one-third the number of observations.

Examining heterogeneity at the school level, some positive effects of pre-K re-emerge. Specifically, in quality states (but not other states), pre-K is estimated to increase reading scores and reduce special education assignments in majority-black schools, to increase reading scores in majority-Hispanic schools, and to increase math scores in heavily white schools.

RESULTS AT 8TH GRADE

We also conducted some estimation of 8th grade outcomes, and found results broadly similar to those from 4th grade.

Table E11 reports effects of pre-K on 8th grade outcomes at the state level. This table is similar to Table 4 and Appendix Table E1, but shows 8th grade rather than 4th grade outcomes.

Table E11: The Effects of Pre-K on State-Level 8th Grade Outcomes: FRL vs Non-FRL students

	(1)	(2)	(3)	(4)
	FRL students		Non-FRL students	
Math scores (percentile)	3.079	0.216	1.168	−0.218
	(2.397)	(2.598)	(2.741)	(3.004)
Reading scores (percentile)	0.106	−0.135	−1.689	−1.570
	(2.467)	(2.167)	(2.930)	(2.138)
Special education (proportion ×100)	0.52	−1.40	−1.24	−2.21
	(2.47)	(2.16)	(1.58)	(1.96)
Over-age for grade (proportion ×100)	−1.69	0.76	−0.26	0.22
	(3.36)	(4.51)	(1.96)	(2.56)
Include state-specific linear time trends	No	Yes	No	Yes

NOTE: Each cell is from a separate regression of the NAEP outcome on the pre-K measure, a set of state dummies, a set of test year dummies, and a quartic in cost-of-living adjusted current spending per student, averaged over the test year and preceding eight years to account for time since pre-K. Each observation is a state-year, and there are 334 for math, 352 for reading, 384 for special ed., and 348 for over-age for grade. Standard errors in parentheses are clustered by state. The underlying dependent variables are state-year cell means that have been regression-adjusted for individual student characteristics and recentered so that the national weighted mean is zero for each year; see text for details. Column (1) shows results for students who are eligible for free or reduced-price lunch (FRL; family income below 185% of poverty line) and column (2) shows results for ineligible students, where eligibility is taken from the NAEP student-level data. The independent variable is the ratio of pre-K enrollment in that state-year to first grade enrollment in that same state-year, taken from the CCD. The coefficients thus reflect the estimated effect of moving from 0 to 100 percent enrollment in pre-K.

Table E11 is similar in its implications to the previous results at the 4th grade level: none of the estimated effects of pre-K is statistically significant, but the estimates are too imprecise to be useful for policy purposes.

Table E12 turns to the district-level data, and reports effects of pre-K on 8th grade outcomes assuming homogeneous treatment. Table E12 is thus similar to Table 5, but at the 8th grade level rather than the 4th grade level.

Table E12: The Effects of Pre-K on District-Level 8th Grade Outcomes

	(1)	(2)
Math scores (percentile)	0.416	0.332
	(0.696)	(0.605)
Reading scores (percentile)	−1.788***	−1.675**
	(0.659)	(0.651)
Special education (proportion ×100)	−0.34	−0.33
	(0.54)	(0.51)
Over-age for grade (proportion ×100)	0.13	0.14
	(0.51)	(0.52)
Include district fixed effects	Yes	Yes
Include district time-varying controls?	No	Yes

NOTE: * significant at the 0.10 level; ** significant at the 0.05 level; *** significant at the 0.01 level. Each cell is from a separate regression of the outcome on the pre-K measure, a set of test year dummies, a quartic in cost-of-living-adjusted, district-level current spending per student (averaged over the test year and preceding four years to account for time since pre-K), and the other controls as shown. District time-varying controls include categorical dummies for the share of students eligible for free or reduced-price lunch, the student enrollment (size) of the district, the share of instructional staff working part-time, the share of students who are black in the district, the share of students who are Hispanic in the district, the number of private district pre-K slots available within 5 km of the district (normalized by the district's grade 1 enrollment), and the number of Head Start four-year-old slots available within 10 km of the district (normalized by the district's grade 1 enrollment). Each observation is a district-year, and there are 18,260 observations (4,880 unique districts) for math scores; 18,690 (4,950) for reading scores; 20,600 (5,210) for special education; and 18,540 (4,910) for over-age for grade. All observation and district counts have been rounded to the nearest 10 to comply with disclosure restrictions. Standard errors in parentheses are clustered by district. The underlying dependent variables are district-year cell means that have been regression-adjusted for individual student characteristics and recentered so that the national weighted mean is zero for each year; see text for details. The independent variable is the ratio of pre-K enrollment in that district-year to first grade enrollment in that same district-year, taken from the CCD. The coefficients thus reflect the estimated effect of moving from 0 to 100 percent enrollment in pre-K.

As with Table 5, when all districts are grouped together, there is no evidence that outcomes are positively affected by greater district pre-K enrollment rates.

Table E13 considers pre-K effects on 8th grade outcomes, but allows effects in "quality states" to differ from other states. Table E13 is similar to text Table 6, but for 8th grade rather than 4th grade outcomes.

Table E13: The Effects of Pre-K on District-Level 8th Grade Outcomes, Quality States vs. Other States

	Other states	Quality states	Difference
Math scores (percentile)	−0.577	3.585***	4.162***
	(0.688)	(1.324)	(1.501)
Reading scores (percentile)	−1.938**	−0.804	1.134
	(0.787)	(0.999)	(1.261)
Special education (proportion ×100)	−0.82	1.30	2.12**
	(0.59)	(0.89)	(1.05)
Over-age for grade (proportion ×100)	0.56	−1.32	−1.87
	(0.60)	(0.97)	(1.14)
Include district fixed effects		Yes	
Include district time-varying controls?		Yes	

NOTE: * significant at the 0.10 level; ** significant at the 0.05 level; *** significant at the 0.01 level. Each row is from a separate regression of the outcome on the pre-K measure, a set of test year dummies, a quartic in cost-of-living-adjusted, district-level current spending per student (averaged over the test year and preceding four years to account for time since pre-K), and district time-varying controls (see note to Appendix Table E12). The coefficients across columns show the pre-K measure, its interaction with an indicator variable for being a "quality program state" (equal to 1 for MD, MA, NJ, NC, and OK), and the net effect of pre-K in quality states. The estimates in column (1) thus shows the impact of moving from 0 to 100 percent enrollment in pre-K in all but the "quality states;" the estimates in column (2) shows the impact of moving from 0 to 100 percent enrollment in pre-K in the "quality states;" and the estimates in column (3) show the difference in these impacts. Each observation is a district-year; for sample sizes, see note to Appendix Table E12.

The results are similar in their implications to Table 6. Pre-K is estimated to have statistically significant and substantively large effects on math test scores in "quality states" but not in other states.